(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 February 2003 (27.02.2003)

PCT

(10) International Publication Number
WO 03/016493 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/26323

(22) International Filing Date: 16 August 2002 (16.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/313,242	17 August 2001 (17.08.2001)	US
60/324,782	21 September 2001 (21.09.2001)	US
60/328,184	2 October 2001 (02.10.2001)	US
60/345,937	26 October 2001 (26.10.2001)	US
60/335,698	1 November 2001 (01.11.2001)	US
60/332,804	13 November 2001 (13.11.2001)	US
60/333,922	27 November 2001 (27.11.2001)	US
60/375,637	26 April 2002 (26.04.2002)	US
60/377,444	3 May 2002 (03.05.2002)	US
60/388,180	11 June 2002 (11.06.2002)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **NGUYEN, Danniel, B.** [US/US]; 4922 Arnica Court, San Jose, CA 95111 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **GREENE, Barrie, D.** [US/US]; 1332 10th Avenue # 104, San Francisco, CA 94122 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **LEE, Sally** [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **GORVAD, Ann, E.** [US/US]; 369 Marie Common, Livermore, CA 94550 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **YUE, Huibin** [US/US]; 1170 South Stelling Road, Cupertino, CA 95014 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **LEHR-MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95037 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **LEE, Soo, Yeun** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **FARIS, Mary** [US/US]; 2538 Almaden Court, Los Angeles, CA 90077 (US). **TURNER, Christopher** [US/US]; 817 Santa Fe Avenue, Stanford, CA 94305 (US). **FURNESS, Michael** [GB/GB]; 2 Brookside, Exning, New Market, Antrim CB8 7EL (GB). **BUCHBINDER, Jenny, L.** [US/US]; 776 Fifth Avenue #1, San Francisco, CA 94118 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A # 205, San Leandro, CA 94577 (US). **LI, Joana, X.** [US/US]; 1264 Geneva Avenue, San Francisco, CA 94112 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **HAFALIA, April, J.A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **LINDQUIST, Erika, A.** [US/US]; 2394 Mariner Square Drive #C-121, Alameda, CA 94501 (US). **JIANG, Xin** [US/US]; 14371 Elva Avenue, Saratoga, CA 95070 (US). **JACKSON, Alan, A.** [US/US]; 1541 Elwood Drive, Los Gatos, CA 95032 (US). **WILSON, Amy, D.** [US/US]; 1056 Continentals Way, #27, Belmont, CA 94002 (US). **JIN, Pei** [US/US]; 320 Curtner Avenue #D, Palo Alto, CA 94306 (US). **KHARE, Reena** [US/US]; 12650 Orella Court, Saratoga, CA 95070 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

[Continued on next page]

(54) Title: TRANSPORTERS AND ION CHANNELS

(57) Abstract: Various embodiments of the invention provide human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.

WO 03/016493 A2

MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TRANSPORTERS AND ION CHANNELS

TECHNICAL FIELD

The invention relates to novel nucleic acids, transporters and ion channels encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of transport, neurological, muscle, immunological and cell proliferative disorders. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and transporters and ion channels.

BACKGROUND OF THE INVENTION

Eukaryotic cells are surrounded and subdivided into functionally distinct organelles by hydrophobic lipid bilayer membranes which are highly impermeable to most polar molecules. Cells and organelles require transport proteins to import and export essential nutrients and metal ions including K^+, NH_4^+, P_i, SO_4^{2-}, sugars, and vitamins, as well as various metabolic waste products. Transport proteins also play roles in antibiotic resistance, toxin secretion, ion balance, synaptic neurotransmission, kidney function, intestinal absorption, tumor growth, and other diverse cell functions (Griffith, J. and C. Sansom (1998) The Transporter Facts Book, Academic Press, San Diego CA, pp. 3-29). Transport can occur by a passive concentration-dependent mechanism, or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Proteins that function in transport include carrier proteins, which bind to a specific solute and undergo a conformational change that translocates the bound solute across the membrane, and channel proteins, which form hydrophilic pores that allow specific solutes to diffuse through the membrane down an electrochemical solute gradient.

Carrier proteins which transport a single solute from one side of the membrane to the other are called uniporters. In contrast, coupled transporters link the transfer of one solute with simultaneous or sequential transfer of a second solute, either in the same direction (symport) or in the opposite direction (antiport). For example, intestinal and kidney epithelium contains a variety of symporter systems driven by the sodium gradient that exists across the plasma membrane. Sodium moves into the cell down its electrochemical gradient and brings the solute into the cell with it. The sodium gradient that provides the driving force for solute uptake is maintained by the ubiquitous Na^+/K^+ ATPase system. Sodium-coupled transporters include the mammalian glucose transporter (SGLT1), iodide transporter (NIS), and multivitamin transporter (SMVT). All three transporters have twelve putative transmembrane segments, extracellular glycosylation sites, and cytoplasmically-oriented N- and C-termini. NIS plays a crucial role in the evaluation, diagnosis, and treatment of

various thyroid pathologies because it is the molecular basis for radioiodide thyroid-imaging techniques and for specific targeting of radioisotopes to the thyroid gland (Levy, O. et al. (1997) Proc. Natl. Acad. Sci. USA 94:5568-5573). SMVT is expressed in the intestinal mucosa, kidney, and placenta, and is implicated in the transport of the water-soluble vitamins, e.g., biotin and pantothenate (Prasad, P.D. et al. (1998) J. Biol. Chem. 273:7501-7506).

One of the largest families of transporters is the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family. MFS transporters are single polypeptide carriers that transport small solutes in response to ion gradients. Members of the MFS are found in all classes of living organisms, and include transporters for sugars, oligosaccharides, phosphates, nitrates, nucleosides, monocarboxylates, and drugs. MFS transporters found in eukaryotes all have a structure comprising 12 transmembrane segments (Pao, S.S. et al. (1998) Microbiol. Molec. Biol. Rev. 62:1-34). The largest family of MFS transporters is the sugar transporter family, which includes the seven glucose transporters (GLUT1-GLUT7) found in humans that are required for the transport of glucose and other hexose sugars. These glucose transport proteins have unique tissue distributions and physiological functions. GLUT1 provides many cell types with their basal glucose requirements and transports glucose across epithelial and endothelial barrier tissues; GLUT2 facilitates glucose uptake or efflux from the liver; GLUT3 regulates glucose supply to neurons; GLUT4 is responsible for insulin-regulated glucose disposal; and GLUT5 regulates fructose uptake into skeletal muscle. Defects in glucose transporters are involved in a recently identified neurological syndrome causing infantile seizures and developmental delay, as well as glycogen storage disease, Fanconi-Bickel syndrome, and non-insulin-dependent diabetes mellitus (Mueckler, M. (1994) Eur. J. Biochem. 219:713-725; Longo, N. and L.J. Elsas (1998) Adv. Pediatr. 45:293-313).

Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H^+-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H^+-linked monocarboxylate transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K_m values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na^+-monocarboxylate

cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are specific and selective transporters for organic cations and organic anions in organs including the kidney, intestine and liver. Organic anion transporters are selective for hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintenance of intercellular pH (Poole, R.C. and A.P. Halestrap (1993) Am. J. Physiol. 264:C761-C782; Price, N.T. et al. (1998) Biochem. J. 329:321-328; and Martinelle, K. and I. Haggstrom (1993) J. Biotechnol. 30:339-350).

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport substances ranging from small molecules such as ions, sugars, amino acids, peptides, and phospholipids, to lipopeptides, large proteins, and complex hydrophobic drugs. ABC transporters consist of four modules: two nucleotide-binding domains (NBD), which hydrolyze ATP to supply the energy required for transport, and two membrane-spanning domains (MSD), each containing six putative transmembrane segments. These four modules may be encoded by a single gene, as is the case for the cystic fibrosis transmembrane regulator (CFTR), or by separate genes. When encoded by separate genes, each gene product contains a single NBD and MSD. These "half-molecules" form homo- and heterodimers, such as Tap1 and Tap2, the endoplasmic reticulum-based major histocompatibility (MHC) peptide transport system. Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP), Zellweger syndrome (peroxisomal membrane protein-70, PMP70), and hyperinsulinemic hypoglycemia (sulfonylurea receptor, SUR). Overexpression of the multidrug resistance (MDR) protein, another ABC transporter, in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and S. Michaelis (1998) Meth. Enzymol. 292:130-162).

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum, selenium, nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and other target organs, where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states (Danks, D.M. (1986) J. Med. Genet. 23:99-106).

Transport of fatty acids across the plasma membrane can occur by diffusion, a high capacity, low affinity process. However, under normal physiological conditions a significant fraction of fatty acid transport appears to occur via a high affinity, low capacity protein-mediated transport process. Fatty acid transport protein (FATP), an integral membrane protein with four transmembrane segments, is expressed in tissues exhibiting high levels of plasma membrane fatty acid flux, such as muscle, heart, and adipose. Expression of FATP is upregulated in 3T3-L1 cells during adipose conversion, and expression in COS7 fibroblasts elevates uptake of long-chain fatty acids (Hui, T.Y. et al. (1998) J. Biol. Chem. 273:27420-27429).

The lipocalin superfamily constitutes a phylogenetically conserved group of more than forty proteins that function as extracellular ligand-binding proteins which bind and transport small hydrophobic molecules. Members of this family function as carriers of retinoids, odorants, chromophores, pheromones, allergens, and sterols, and in a variety of processes including nutrient transport, cell growth regulation, immune response, and prostaglandin synthesis. A subset of these proteins may be multifunctional, serving as either a biosynthetic enzyme or as a specific enzyme inhibitor. (Tanaka, T. et al. (1997) J. Biol. Chem. 272:15789-15795; and van't Hof, W. et al. (1997) J. Biol. Chem. 272:1837-1841.)

Members of the lipocalin family display unusually low levels of overall sequence conservation. Pairwise sequence identity often falls below 20%. Sequence similarity between family members is limited to conserved cysteines which form disulfide bonds and three motifs which form a juxtaposed cluster that functions as a target cell recognition site. The lipocalins share an eight stranded, anti-parallel beta-sheet which folds back on itself to form a continuously hydrogen-bonded beta-barrel. The pocket formed by the barrel functions as an internal ligand binding site. Seven loops (L1 to L7) form short beta-hairpins, except loop L1 which is a large omega loop that forms a lid to partially close the internal ligand-binding site (Flower (1996) Biochem. J. 318:1-14).

Lipocalins are important transport molecules. Each lipocalin associates with a particular ligand and delivers that ligand to appropriate target sites within the organism. Retinol-binding protein (RBP), one of the best characterized lipocalins, transports retinol from stores within the liver to target tissues. Apolipoprotein D (apo D), a component of high density lipoproteins (HDLs) and low density lipoproteins (LDLs), functions in the targeted collection and delivery of cholesterol throughout the body. Lipocalins are also involved in cell regulatory processes. Apo D, which is identical to gross-cystic-disease-fluid protein (GCDFP)-24, is a progesterone/pregnenolone-binding protein expressed at high levels in breast cyst fluid. Secretion of apo D in certain human breast cancer cell lines is accompanied by reduced cell proliferation and progression of cells to a more differentiated phenotype.

Similarly, apo D and another lipocalin, α_1-acid glycoprotein (AGP), are involved in nerve cell regeneration. AGP is also involved in anti-inflammatory and immunosuppressive activities. AGP is one of the positive acute-phase proteins (APP); circulating levels of AGP increase in response to stress and inflammatory stimulation. AGP accumulates at sites of inflammation where it inhibits platelet and neutrophil activation and inhibits phagocytosis. The immunomodulatory properties of AGP are due to glycosylation. AGP is 40% carbohydrate, making it unusually acidic and soluble. The glycosylation pattern of AGP changes during acute-phase response, and deglycosylated AGP has no immunosuppressive activity (Flower (1994) FEBS Lett. 354:7-11; Flower (1996) *supra*).

The lipocalin superfamily also includes several animal allergens, including the mouse major urinary protein (mMUP), the rat α-2-microgloobulin (rA2U), the bovine β-lactoglobulin (βlg), the cockroach allergen (Bla g4), bovine dander allergen (Bos d2), and the major horse allergen, designated *Equus caballus* allergen 1 (Equ c1). Equ c1 is a powerful allergen responsible for about 80% of anti-horse IgE antibody response in patients who are chronically exposed to horse allergens. It appears that lipocalins may contain a common structure that is able to induce the IgE response (Gregoire, C. et al., (1996) J. Biol. Chem. 271:32951-32959).

Lipocalins are used as diagnostic and prognostic markers in a variety of disease states. The plasma level of AGP is monitored during pregnancy and in diagnosis and prognosis of conditions including cancer chemotherapy, renal disfunction, myocardial infarction, arthritis, and multiple sclerosis. RBP is used clinically as a marker of tubular reabsorption in the kidney, and apo D is a marker in gross cystic breast disease (Flower (1996) *supra*). Additionally, the use of lipocalin animal allergens may help in the diagnosis of allergic reactions to horses (Gregoire *supra*), pigs, cockroaches, mice and rats.

Mitochondrial carrier proteins are transmembrane-spanning proteins which transport ions and charged metabolites between the cytosol and the mitochondrial matrix. Examples include the ADP, ATP carrier protein; the 2-oxoglutarate/malate carrier; the phosphate carrier protein; the pyruvate carrier; the dicarboxylate carrier which transports malate, succinate, fumarate, and phosphate; the tricarboxylate carrier which transports citrate and malate; and the Grave's disease carrier protein, a protein recognized by IgG in patients with active Grave's disease, an autoimmune disorder resulting in hyperthyroidism. Proteins in this family consist of three tandem repeats of an approximately 100 amino acid domain, each of which contains two transmembrane regions (Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, p. 551; PROSITE PDOC00189 Mitochondrial energy transfer proteins signature; Online Mendelian Inheritance in Man (OMIM) *275000 Graves Disease).

This class of transporters also includes the mitochondrial uncoupling proteins, which create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. The result is energy dissipation in the form of heat. Mitochondrial uncoupling proteins have been implicated as modulators of thermoregulation and metabolic rate, and have been proposed as potential targets for drugs against metabolic diseases such as obesity (Ricquier, D. et al. (1999) J. Int. Med. 245:637-642).

Ion Channels

The electrical potential of a cell is generated and maintained by controlling the movement of ions across the plasma membrane. The movement of ions requires ion channels, which form ion-selective pores within the membrane. There are two basic types of ion channels, ion transporters and gated ion channels. Ion transporters utilize the energy obtained from ATP hydrolysis to actively transport an ion against the ion's concentration gradient. Gated ion channels allow passive flow of an ion down the ion's electrochemical gradient under restricted conditions. Together, these types of ion channels generate, maintain, and utilize an electrochemical gradient that is used in 1) electrical impulse conduction down the axon of a nerve cell, 2) transport of molecules into cells against concentration gradients, 3) initiation of muscle contraction, and 4) endocrine cell secretion.

Ion Transporters

Ion transporters generate and maintain the resting electrical potential of a cell. Utilizing the energy derived from ATP hydrolysis, they transport ions against the ion's concentration gradient. These transmembrane ATPases are divided into three families. The phosphorylated (P) class ion transporters, including Na^+-K^+ ATPase, Ca^{2+}-ATPase, and H^+-ATPase, are activated by a phosphorylation event. P-class ion transporters are responsible for maintaining resting potential distributions such that cytosolic concentrations of Na^+ and Ca^{2+} are low and cytosolic concentration of K^+ is high. The vacuolar (V) class of ion transporters includes H^+ pumps on intracellular organelles, such as lysosomes and Golgi. V-class ion transporters are responsible for generating the low pH within the lumen of these organelles that is required for function. The coupling factor (F) class consists of H^+ pumps in the mitochondria. F-class ion transporters utilize a proton gradient to generate ATP from ADP and inorganic phosphate (P_i).

The P-ATPases are hexamers of a 100 kD subunit with ten transmembrane domains and several large cytoplasmic regions that may play a role in ion binding (Scarborough, G.A. (1999) Curr. Opin. Cell Biol. 11:517-522). The V-ATPases are composed of two functional domains: the V_1 domain, a peripheral complex responsible for ATP hydrolysis; and the V_0 domain, an integral complex responsible for proton translocation across the membrane. The F-ATPases are structurally and

evolutionarily related to the V-ATPases. The F-ATPase F_0 domain contains 12 copies of the c subunit, a highly hydrophobic protein composed of two transmembrane domains and containing a single buried carboxyl group in TM2 that is essential for proton transport. The V-ATPase V_0 domain contains three types of homologous c subunits with four or five transmembrane domains and the essential carboxyl group in TM4 or TM3. Both types of complex also contain a single a subunit that may be involved in regulating the pH dependence of activity (Forgac, M. (1999) J. Biol. Chem. 274:12951-12954).

The resting potential of the cell is utilized in many processes involving carrier proteins and gated ion channels. Carrier proteins utilize the resting potential to transport molecules into and out of the cell. Amino acid and glucose transport into many cells is linked to sodium ion co-transport (symport) so that the movement of Na^+ down an electrochemical gradient drives transport of the other molecule up a concentration gradient. Similarly, cardiac muscle links transfer of Ca^{2+} out of the cell with transport of Na^+ into the cell (antiport).

Gated Ion Channels

Gated ion channels control ion flow by regulating the opening and closing of pores. The ability to control ion flux through various gating mechanisms allows ion channels to mediate such diverse signaling and homeostatic functions as neuronal and endocrine signaling, muscle contraction, fertilization, and regulation of ion and pH balance. Gated ion channels are categorized according to the manner of regulating the gating function. Mechanically-gated channels open their pores in response to mechanical stress; voltage-gated channels (e.g., Na^+, K^+, Ca^{2+}, and Cl^- channels) open their pores in response to changes in membrane potential; and ligand-gated channels (e.g., acetylcholine-, serotonin-, and glutamate-gated cation channels, and GABA- and glycine-gated chloride channels) open their pores in the presence of a specific ion, nucleotide, or neurotransmitter. The gating properties of a particular ion channel (i.e., its threshold for and duration of opening and closing) are sometimes modulated by association with auxiliary channel proteins and/or post translational modifications, such as phosphorylation.

Mechanically-gated or mechanosensitive ion channels act as transducers for the senses of touch, hearing, and balance, and also play important roles in cell volume regulation, smooth muscle contraction, and cardiac rhythm generation. A stretch-inactivated channel (SIC) was recently cloned from rat kidney. The SIC channel belongs to a group of channels which are activated by pressure or stress on the cell membrane and conduct both Ca^{2+} and Na^+ (Suzuki, M. et al. (1999) J. Biol. Chem. 274:6330-6335).

The pore-forming subunits of the voltage-gated cation channels form a superfamily of ion

channel proteins. The characteristic domain of these channel proteins comprises six transmembrane domains (S1-S6), a pore-forming region (P) located between S5 and S6, and intracellular amino and carboxy termini. In the Na^+ and Ca^{2+} subfamilies, this domain is repeated four times, while in the K^+ channel subfamily, each channel is formed from a tetramer of either identical or dissimilar subunits. The P region contains information specifying the ion selectivity for the channel. In the case of K^+ channels, a GYG tripeptide is involved in this selectivity (Ishii, T.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11651-11656).

Voltage-gated Na^+ and K^+ channels are necessary for the function of electrically excitable cells, such as nerve and muscle cells. Action potentials, which lead to neurotransmitter release and muscle contraction, arise from large, transient changes in the permeability of the membrane to Na^+ and K^+ ions. Depolarization of the membrane beyond the threshold level opens voltage-gated Na^+ channels. Sodium ions flow into the cell, further depolarizing the membrane and opening more voltage-gated Na^+ channels, which propagates the depolarization down the length of the cell. Depolarization also opens voltage-gated potassium channels. Consequently, potassium ions flow outward, which leads to repolarization of the membrane. Voltage-gated channels utilize charged residues in the fourth transmembrane segment (S4) to sense voltage change. The open state lasts only about 1 millisecond, at which time the channel spontaneously converts into an inactive state that cannot be opened irrespective of the membrane potential. Inactivation is mediated by the channel's N-terminus, which acts as a plug that closes the pore. The transition from an inactive to a closed state requires a return to resting potential.

Voltage-gated Na^+ channels are heterotrimeric complexes composed of a 260 kDa pore-forming α subunit that associates with two smaller auxiliary subunits, $\beta 1$ and $\beta 2$. The $\beta 2$ subunit is a integral membrane glycoprotein that contains an extracellular Ig domain, and its association with α and $\beta 1$ subunits correlates with increased functional expression of the channel, a change in its gating properties, as well as an increase in whole cell capacitance due to an increase in membrane surface area (Isom, L.L. et al. (1995) Cell 83:433-442).

Non voltage-gated Na^+ channels include the members of the amiloride-sensitive Na^+ channel/degenerin (NaC/DEG) family. Channel subunits of this family are thought to consist of two transmembrane domains flanking a long extracellular loop, with the amino and carboxyl termini located within the cell. The NaC/DEG family includes the epithelial Na^+ channel (ENaC) involved in Na^+ reabsorption in epithelia including the airway, distal colon, cortical collecting duct of the kidney, and exocrine duct glands. Mutations in ENaC result in pseudohypoaldosteronism type 1 and Liddle's syndrome (pseudohyperaldosteronism). The NaC/DEG family also includes the recently characterized

H^+-gated cation channels or acid-sensing ion channels (ASIC). ASIC subunits are expressed in the brain and form heteromultimeric Na^+-permeable channels. These channels require acid pH fluctuations for activation. ASIC subunits show homology to the degenerins, a family of mechanically-gated channels originally isolated from C. elegans. Mutations in the degenerins cause neurodegeneration. ASIC subunits may also have a role in neuronal function, or in pain perception, since tissue acidosis causes pain (Waldmann, R. and M. Lazdunski (1998) Curr. Opin. Neurobiol. 8:418-424; Eglen, R.M. et al. (1999) Trends Pharmacol. Sci. 20:337-342).

K^+ channels are located in all cell types, and may be regulated by voltage, ATP concentration, or second messengers such as Ca^{2+} and cAMP. In non-excitable tissue, K^+ channels are involved in protein synthesis, control of endocrine secretions, and the maintenance of osmotic equilibrium across membranes. In neurons and other excitable cells, in addition to regulating action potentials and repolarizing membranes, K^+ channels are responsible for setting the resting membrane potential. The cytosol contains non-diffusible anions and, to balance this net negative charge, the cell contains a Na^+-K^+ pump and ion channels that provide the redistribution of Na^+, K^+, and Cl^-. The pump actively transports Na^+ out of the cell and K^+ into the cell in a 3:2 ratio. Ion channels in the plasma membrane allow K^+ and Cl^- to flow by passive diffusion. Because of the high negative charge within the cytosol, Cl^- flows out of the cell. The flow of K^+ is balanced by an electromotive force pulling K^+ into the cell, and a K^+ concentration gradient pushing K^+ out of the cell. Thus, the resting membrane potential is primarily regulated by K^+ flow (Salkoff, L. and T. Jegla (1995) Neuron 15:489-492).

Potassium channel subunits of the *Shaker*-like superfamily all have the characteristic six transmembrane/1 pore domain structure. Four subunits combine as homo- or heterotetramers to form functional K channels. These pore-forming subunits also associate with various cytoplasmic β subunits that alter channel inactivation kinetics. The *Shaker*-like channel family includes the voltage-gated K^+ channels as well as the delayed rectifier type channels such as the human ether-a-go-go related gene (HERG) associated with long QT, a cardiac dysrythmia syndrome (Curran, M.E. (1998) Curr. Opin. Biotechnol. 9:565-572; Kaczorowski, G.J. and M.L. Garcia (1999) Curr. Opin. Chem. Biol. 3:448-458).

A second superfamily of K^+ channels is composed of the inward rectifying channels (Kir). Kir channels have the property of preferentially conducting K^+ currents in the inward direction. These proteins consist of a single potassium selective pore domain and two transmembrane domains, which correspond to the fifth and sixth transmembrane domains of voltage-gated K^+ channels. Kir subunits also associate as tetramers. The Kir family includes ROMK1, mutations in which lead to Bartter syndrome, a renal tubular disorder. Kir channels are also involved in regulation of cardiac pacemaker

activity, seizures and epilepsy, and insulin regulation (Doupnik, C.A. et al. (1995) Curr. Opin. Neurobiol. 5:268-277; Curran, *supra*).

The recently recognized TWIK K^+ channel family includes the mammalian TWIK-1, TREK-1 and TASK proteins. Members of this family possess an overall structure with four transmembrane domains and two P domains. These proteins are probably involved in controlling the resting potential in a large set of cell types (Duprat, F. et al. (1997) EMBO J 16:5464-5471).

The voltage-gated Ca^{2+} channels have been classified into several subtypes based upon their electrophysiological and pharmacological characteristics. L-type Ca^{2+} channels are predominantly expressed in heart and skeletal muscle where they play an essential role in excitation-contraction coupling. T-type channels are important for cardiac pacemaker activity, while N-type and P/Q-type channels are involved in the control of neurotransmitter release in the central and peripheral nervous system. The L-type and N-type voltage-gated Ca^{2+} channels have been purified and, though their functions differ dramatically, they have similar subunit compositions. The channels are composed of three subunits. The α_1 subunit forms the membrane pore and voltage sensor, while the $\alpha_2\delta$ and β subunits modulate the voltage-dependence, gating properties, and the current amplitude of the channel. These subunits are encoded by at least six α_1, one $\alpha_2\delta$, and four β genes. A fourth subunit, γ, has been identified in skeletal muscle (Walker, D. et al. (1998) J. Biol. Chem. 273:2361-2367; McCleskey, E.W. (1994) Curr. Opin. Neurobiol. 4:304-312).

The high-voltage-activated Ca^{2+} channels that have been characterized biochemically include complexes of a pore-forming alpha1 subunit of approximately 190-250 kDa; a transmembrane complex of alpha2 and delta subunits; an intracellular beta subunit; and in some cases a transmembrane gamma subunit. A variety of alpha1 subunits, alpha2delta complexes, beta subunits, and gamma subunits are known. The Cav1 family of alpha1 subunits conduct L-type Ca^{2+} currents, which initiate muscle contraction, endocrine secretion, and gene transcription, and are regulated primarily by second messenger-activated protein phosphorylation pathways. The Cav2 family of alpha1 subunits conduct N-type, P/Q-type, and R-type Ca^{2+} currents, which initiate rapid synaptic transmission and are regulated primarily by direct interaction with G proteins and SNARE proteins and secondarily by protein phosphorylation. The Cav3 family of alpha1 subunits conduct T-type Ca^{2+} currents, which are activated and inactivated more rapidly and at more negative membrane potentials than other Ca^{2+} current types. The distinct structures and patterns of regulation of these three families of Ca^{2+} channels provide an array of Ca^{2+} entry pathways in response to changes in membrane potential and a range of possibilities for regulation of Ca^{2+} entry by second messenger pathways and interacting proteins (Catterall, W.A. (2000) Annu. Rev. Cell Dev. Biol. 16:521-555).

The alpha-2 subunit of the voltage-gated Ca^{2+}-channel may include one or more Cache domains. An extracellular Cache domain may be fused to an intracellular catalytic domain, such as the histidine kinase, PP2C phosphatase, GGDEF (a predicted diguanylate cyclase), HD-GYP (a predicted phosphodiesterase) or adenylyl cyclase domain, or to a noncatalytic domain, like the methyl-accepting, DNA-binding winged helix-turn-helix, GAF, PAS or HAMP (a domain found in istidine kinases, denylyl cyclases, ethyl-binding proteins and phosphatases). Small molecules are bound via the Cache domain and this signal is converted into diverse outputs depending on the intracellular domains (Anantharaman, V. and Aravind, L.(2000) Trends Biochem. Sci. 25:535-537).

The transient receptor family (Trp) of calcium ion channels are thought to mediate capacitative calcium entry (CCE). CCE is the Ca^{2+} influx into cells to resupply Ca^{2+} stores depleted by the action of inositol triphosphate (IP3) and other agents in response to numerous hormones and growth factors. Trp and Trp-like were first cloned from Drosophila and have similarity to voltage gated Ca^{2+} channels in the S3 through S6 regions. This suggests that Trp and/or related proteins may form mammalian CCE channels (Zhu, X. et al. (1996) Cell 85:661-671; Boulay, G. et al. (1997) J. Biol. Chem. 272:29672-29680). Melastatin is a gene isolated in both the mouse and human, whose expression in melanoma cells is inversely correlated with melanoma aggressiveness *in vivo*. The human cDNA transcript corresponds to a 1533-amino acid protein having homology to members of the Trp family. It has been proposed that the combined use of malastatin mRNA expression status and tumor thickness might allow for the determination of subgroups of patients at both low and high risk for developing metastatic disease (Duncan, L.M. et al (2001) J. Clin. Oncol. 19:568-576).

Chloride channels are necessary in endocrine secretion and in regulation of cytosolic and organelle pH. In secretory epithelial cells, Cl^- enters the cell across a basolateral membrane through an Na^+, K^+/Cl^- cotransporter, accumulating in the cell above its electrochemical equilibrium concentration. Secretion of Cl^- from the apical surface, in response to hormonal stimulation, leads to flow of Na^+ and water into the secretory lumen. The cystic fibrosis transmembrane conductance regulator (CFTR) is a chloride channel encoded by the gene for cystic fibrosis, a common fatal genetic disorder in humans. CFTR is a member of the ABC transporter family, and is composed of two domains each consisting of six transmembrane domains followed by a nucleotide-binding site. Loss of CFTR function decreases transepithelial water secretion and, as a result, the layers of mucus that coat the respiratory tree, pancreatic ducts, and intestine are dehydrated and difficult to clear. The resulting blockage of these sites leads to pancreatic insufficiency, "meconium ileus", and devastating "chronic obstructive pulmonary disease" (Al-Awqati, Q. et al. (1992) J. Exp. Biol. 172:245-266).

The voltage-gated chloride channels (CLC) are characterized by 10-12 transmembrane

domains, as well as two small globular domains known as CBS domains. The CLC subunits probably function as homotetramers. CLC proteins are involved in regulation of cell volume, membrane potential stabilization, signal transduction, and transepithelial transport. Mutations in CLC-1, expressed predominantly in skeletal muscle, are responsible for autosomal recessive generalized myotonia and autosomal dominant myotonia congenita, while mutations in the kidney channel CLC-5 lead to kidney stones (Jentsch, T.J. (1996) Curr. Opin. Neurobiol. 6:303-310).

Ligand-gated channels open their pores when an extracellular or intracellular mediator binds to the channel. Neurotransmitter-gated channels are channels that open when a neurotransmitter binds to their extracellular domain. These channels exist in the postsynaptic membrane of nerve or muscle cells. There are two types of neurotransmitter-gated channels. Sodium channels open in response to excitatory neurotransmitters, such as acetylcholine, glutamate, and serotonin. This opening causes an influx of Na^+ and produces the initial localized depolarization that activates the voltage-gated channels and starts the action potential. Chloride channels open in response to inhibitory neurotransmitters, such as γ-aminobutyric acid (GABA) and glycine, leading to hyperpolarization of the membrane and the subsequent generation of an action potential. Neurotransmitter-gated ion channels have four transmembrane domains and probably function as pentamers (Jentsch, *supra*). Amino acids in the second transmembrane domain appear to be important in determining channel permeation and selectivity (Sather, W.A. et al. (1994) Curr. Opin. Neurobiol. 4:313-323).

Ligand-gated channels can be regulated by intracellular second messengers. For example, calcium-activated K^+ channels are gated by internal calcium ions. In nerve cells, an influx of calcium during depolarization opens K^+ channels to modulate the magnitude of the action potential (Ishi et al., *supra*). The large conductance (BK) channel has been purified from brain and its subunit composition determined. The α subunit of the BK channel has seven rather than six transmembrane domains in contrast to voltage-gated K^+ channels. The extra transmembrane domain is located at the subunit N-terminus. A 28-amino-acid stretch in the C-terminal region of the subunit (the "calcium bowl" region) contains many negatively charged residues and is thought to be the region responsible for calcium binding. The β subunit consists of two transmembrane domains connected by a glycosylated extracellular loop, with intracellular N- and C-termini (Kaczorowski, *supra*; Vergara, C. et al. (1998) Curr. Opin. Neurobiol. 8:321-329).

Cyclic nucleotide-gated (CNG) channels are gated by cytosolic cyclic nucleotides. The best examples of these are the cAMP-gated Na^+ channels involved in olfaction and the cGMP-gated cation channels involved in vision. Both systems involve ligand-mediated activation of a G-protein coupled receptor which then alters the level of cyclic nucleotide within the cell. CNG channels also

represent a major pathway for Ca^{2+} entry into neurons, and play roles in neuronal development and plasticity. CNG channels are tetramers containing at least two types of subunits, an α subunit which can form functional homomeric channels, and a β subunit, which modulates the channel properties. All CNG subunits have six transmembrane domains and a pore forming region between the fifth and sixth transmembrane domains, similar to voltage-gated K^+ channels. A large C-terminal domain contains a cyclic nucleotide binding domain, while the N-terminal domain confers variation among channel subtypes (Zufall, F. et al. (1997) Curr. Opin. Neurobiol. 7:404-412).

The activity of other types of ion channel proteins may also be modulated by a variety of intracellular signaling proteins. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Kir channels are activated by the binding of the Gβγ subunits of heterotrimeric G-proteins (Reimann, F. and F.M. Ashcroft (1999) Curr. Opin. Cell. Biol. 11:503-508). Other proteins are involved in the localization of ion channels to specific sites in the cell membrane. Such proteins include the PDZ domain proteins known as MAGUKs (membrane-associated guanylate kinases) which regulate the clustering of ion channels at neuronal synapses (Craven, S.E. and D.S. Bredt (1998) Cell 93:495-498).

Disease Correlation

The etiology of numerous human diseases and disorders can be attributed to defects in the transport of molecules across membranes. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g., cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Single-gene defect diseases resulting in an inability to transport small molecules across membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease (van't Hoff, W.G. (1996) Exp. Nephrol. 4:253-262; Talente, G.M. et al. (1994) Ann. Intern. Med. 120:218-226; and Chillon, M. et al. (1995) New Engl. J. Med. 332:1475-1480).

Human diseases caused by mutations in ion channel genes include disorders of skeletal muscle, cardiac muscle, and the central nervous system. Mutations in the pore-forming subunits of sodium and chloride channels cause myotonia, a muscle disorder in which relaxation after voluntary contraction is delayed. Sodium channel myotonias have been treated with channel blockers. Mutations in muscle sodium and calcium channels cause forms of periodic paralysis, while mutations in the sarcoplasmic calcium release channel, T-tubule calcium channel, and muscle sodium channel cause malignant hyperthermia. Cardiac arrythmia disorders such as the long QT syndromes and idiopathic ventricular fibrillation are caused by mutations in potassium and sodium channels (Cooper,

E.C. and L.Y. Jan (1998) Proc. Natl. Acad. Sci. USA 96:4759-4766). All four known human idiopathic epilepsy genes code for ion channel proteins (Berkovic, S.F. and I.E. Scheffer (1999) Curr. Opin. Neurology 12:177-182). Other neurological disorders such as ataxias, hemiplegic migraine and hereditary deafness can also result from mutations in ion channel genes (Jen, J. (1999) Curr. Opin. Neurobiol. 9:274-280; Cooper, *supra*).

Ion channels have been the target for many drug therapies. Neurotransmitter-gated channels have been targeted in therapies for treatment of insomnia, anxiety, depression, and schizophrenia. Voltage-gated channels have been targeted in therapies for arrhythmia, ischemic stroke, head trauma, and neurodegenerative disease (Taylor, C.P. and L.S. Narasimhan (1997) Adv. Pharmacol. 39:47-98). Various classes of ion channels also play an important role in the perception of pain, and thus are potential targets for new analgesics. These include the vanilloid-gated ion channels, which are activated by the vanilloid capsaicin, as well as by noxious heat. Local anesthetics such as lidocaine and mexiletine which blockade voltage-gated Na^+ channels have been useful in the treatment of neuropathic pain (Eglen, *supra*).

Ion channels in the immune system have recently been suggested as targets for immunomodulation. T-cell activation depends upon calcium signaling, and a diverse set of T-cell specific ion channels has been characterized that affect this signaling process. Channel blocking agents can inhibit secretion of lymphokines, cell proliferation, and killing of target cells. A peptide antagonist of the T-cell potassium channel Kv1.3 was found to suppress delayed-type hypersensitivity and allogenic responses in pigs, validating the idea of channel blockers as safe and efficacious immunosuppressants (Cahalan, M.D. and K.G. Chandy (1997) Curr. Opin. Biotechnol. 8:749-756).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling

cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

The potential application of gene expression profiling is relevant to improving the diagnosis, prognosis, and treatment of cancers, such as breast cancer, lung cancer, prostate cancer, ovarian cancer, and bone cancer, as well as the treatment of vascular inflammation and immune responses, liver toxicity, and neurological disorders.

Breast cancer

More than 180,000 new cases of breast cancer are diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, *supra*). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S.S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a

human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down-regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix, W. et al (1999) FEBS Lett 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S.W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Lung cancer

Lung cancer is the leading cause of cancer death in the United States, affecting more than 100,000 men and 50,000 women each year. Nearly 90% of the patients diagnosed with lung cancer are cigarette smokers. Tobacco smoke contains thousands of noxious substances that induce carcinogen metabolizing enzymes and covalent DNA adduct formation in the exposed bronchial epithelium. In nearly 80% of patients diagnosed with lung cancer, metastasis has already occurred. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. The decision to treat with surgery, radiation therapy, or chemotherapy is made on the basis of tumor histology, response to growth factors or hormones, and sensitivity to inhibitors or drugs. With current treatments, most patients die within one year of diagnosis. Earlier diagnosis and a systematic approach to identification, staging, and treatment of lung cancer could positively affect patient outcome.

Lung cancers progress through a series of morphologically distinct stages from hyperplasia to invasive carcinoma. Malignant lung cancers are divided into two groups comprising four histopathological classes. The Non Small Cell Lung Carcinoma (NSCLC) group includes squamous cell carcinomas, adenocarcinomas, and large cell carcinomas and accounts for about 70% of all lung cancer cases. Adenocarcinomas typically arise in the peripheral airways and often form mucin secreting glands. Squamous cell carcinomas typically arise in proximal airways. The histogenesis of

squamous cell carcinomas may be related to chronic inflammation and injury to the bronchial epithelium, leading to squamous metaplasia. The Small Cell Lung Carcinoma (SCLC) group accounts for about 20% of lung cancer cases. SCLCs typically arise in proximal airways and exhibit a number of paraneoplastic syndromes including inappropriate production of adrenocorticotropin and anti-diuretic hormone.

Lung cancer cells accumulate numerous genetic lesions, many of which are associated with cytologically visible chromosomal aberrations. The high frequency of chromosomal deletions associated with lung cancer may reflect the role of multiple tumor suppressor loci in the etiology of this disease. Deletion of the short arm of chromosome 3 is found in over 90% of cases and represents one of the earliest genetic lesions leading to lung cancer. Deletions at chromosome arms 9p and 17p are also common. Other frequently observed genetic lesions include overexpression of telomerase, activation of oncogenes such as K-ras and c-myc, and inactivation of tumor suppressor genes such as RB, p53 and CDKN2.

Genes differentially regulated in lung cancer have been identified by a variety of methods. Using mRNA differential display technology, Manda et al. (1999; Genomics 51:5-14) identified five genes differentially expressed in lung cancer cell lines compared to normal bronchial epithelial cells. Among the known genes, pulmonary surfactant apoprotein A and alpha 2 macroglobulin were down regulated whereas nm23H1 was upregulated. Petersen et al.. (2000; Int J. Cancer, 86:512-517) used suppression subtractive hybridization to identify 552 clones differentially expressed in lung tumor derived cell lines, 205 of which represented known genes. Among the known genes, thrombospondin-1, fibronectin, intercellular adhesion molecule 1, and cytokeratins 6 and 18 were previously observed to be differentially expressed in lung cancers. Wang et al. (2000; Oncogene 19:1519-1528) used a combination of microarray analysis and subtractive hybridization to identify 17 genes differentially overexpresssed in squamous cell carcinoma compared with normal lung epithelium. Among the known genes they identified were keratin isoform 6, KOC, SPRC, IGFb2, connexin 26, plakofillin 1 and cytokeratin 13.

Prostate Cancer

Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic

acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen sensitive and respond to androgen ablation, a population of androgen independent cells evolve from the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. A variety of genes may be differentially expressed during tumor progression. For example, loss of heterozygosity (LOH) is frequently observed on chromosome 8p in prostate cancer. Fluorescence *in situ* hybridization (FISH) revealed a deletion for at least 1 locus on 8p in 29 (69%) tumors, with a significantly higher frequency of the deletion on 8p21.2-p21.1 in advanced prostate cancer than in localized prostate cancer, implying that deletions on 8p22-p21.3 play an important role in tumor differentiation, while 8p21.2-p21.1 deletion plays a role in progression of prostate cancer (Oba, K. et al. (2001) Cancer Genet. Cytogenet. 124: 20-26).

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGFα) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin, J. et al. (1999) Cancer Res. 59:2891-2897; Putz, T. et al. (1999) Cancer Res. 59:227-233). The TGF-β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival

(Gold, L.I. (1999) Crit. Rev. Oncog. 10:303-360). Finally, there are human cell lines representing both the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung, T.D. (1999) Prostate 15:199-207).

Ovarian cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rate for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

Bone cancer

Osteosarcoma is the most common malignant bone tumor in children. Approximately 80% of patients present with non-metastatic disease. After the diagnosis is made by an initial biopsy, treatment involves the use of 3–4 courses of neoadjuvant chemotherapy before definitive surgery, followed by post-operative chemotherapy. With currently available treatment regimens, approximately 30–40% of patients with non-metastatic disease relapse after therapy. Currently, there is no prognostic factor that can be used at the time of initial diagnosis to predict which patients will have a high risk of relapse. The only significant prognostic factor predicting the outcome in a patient with non-metastatic osteosarcoma is the histopathologic response of the primary tumor resected at the time of definitive surgery. The degree of necrosis in the primary tumor is a reflection of the tumor response to neoadjuvant chemotherapy. A higher degree of necrosis (good or favorable response) is associated with a lower risk of relapse and a better outcome. Patients with a lower degree of necrosis (poor or unfavorable response) have a much higher risk of relapse and poor outcome even after complete resection of the primary tumor. Unfortunately, poor outcome cannot be altered despite modification of post-operative chemotherapy to account for the resistance of the primary tumor to neoadjuvant chemotherapy. Thus, there is an urgent need to identify prognostic factors that can be used at the time of diagnosis to recognize the subtypes of osteosarcomas that have various risks of relapse, so that more appropriate chemotherapy can be used at the outset to improve the outcome.

Inflammation and immune responses

Atherosclerosis is a pathological condition characterized by a chronic local inflammatory response within the vessel wall of major arteries. Disease progression results in the formation of

atherosclerotic lesions, unstable plaques which occasionally rupture, precipitating a catastrophic thrombotic occlusion of the vessel lumen. Atherosclerosis and the associated coronary artery disease and cerebral stroke represent the most common causes of death in industrialized nations. Although certain key risk factors have been identified, a full molecular characterization that elucidates the causes and identifies all potential therapeutic targets for this complex disease has not been achieved. Molecular characterization of atherosclerosis requires identification of the genes that contribute to lesion growth, stability, dissolution, rupture and induction of occlusive vessel thrombi.

Blood vessel walls are composed of two tissue layers: an endothelial cell (EC) layer which comprises the lumenal surface of the vessel, and an underlying vascular smooth muscle cell (VSMC) layer. Through dynamic interactions with each other and with surrounding tissues, the vascular endothelium and smooth muscle tissues maintain vascular tone, control selective permeability of the vascular wall, direct vessel remodeling and angiogenesis, and modulate inflammatory and immune responses.

The inflammatory response is a complex vascular reaction mediated by numerous cytokines, chemokines, growth factors, and other signaling molecules expressed by activated ECs, VSMCs and leukocytes. Inflammation protects the organism during trauma and infection, but can also lead to pathological conditions such as atherosclerosis. Activation of vascular endothelium is a central event in a wide range of physiological and disease processes such as vascular tone regulation, coagulation and thrombosis, atherosclerosis, inflammation and some infectious diseases.

The pro-inflammatory cytokines, interleukin (IL)-1 and tumor necrosis factor (TNF), are secreted by a small number of activated macrophages or other cells and can set off a cascade of vascular changes, largely through their ability to alter gene expression patterns in ECs and VSMCs. These vascular changes include vasodilation and increased permeability of microvasculature, edema, and leukocyte extravasation and transmigration across the vessel wall. Ultimately, leukocytes, particularly neutrophils and monocytes/macrophages, accumulate in the extravascular space, where they remove injurious agents by phagocytosis and oxidative killing, a process accompanied by release of toxic factors, such as proteases and reactive oxygen species.

IL-1 and TNF induce pro-inflammatory, thrombotic, and anti-apoptotic changes in gene expression by signaling through receptors on the surface of ECs and VSMCs; these receptors activate transcription factors such as NF*k*B as well as AP-1, IRF-1, and NF-GMa, leading to alterations in gene expression. Genes known to be differentially regulated in EC by IL-1 and TNF include E selectin, VCAM-1, ICAM-1, PAF, I*k*Bα, IAP-1, MCP-1, eotaxin, ENA-78, G-CSF, A20, ICE, and complement C3 component. A key event in inflammation, adhesion and transmigration of blood

leukocytes across the vascular endothelium, for example, is mediated by increased expression of E selectin, P selectin, ICAM-1, and VCAM-1 on activated endothelium.

Several investigators have examined changes in vascular cell gene expression associated with various inflammatory diseases or model systems. Examining human umbilical vein endothelial cells (HUVEC) activated by recombinant TNF-α or conditioned medium from activated human primary monocytes, Horrevoets et al. (1999; Blood 93:3418-3431) identified 106 differentially regulated genes. In a similar approach, deVries et al. (2000; JBC 275:23939-23947) identified 40 differentially regulated genes in umbilical cord artery-derived smooth muscle cells activated by conditioned media from cultured macrophages after stimulation with oxidized LDL particles. In both studies, many of the identified genes were already known to be involved in inflammation. Comparing expression profiles from inflammatory diseased tissues, cultured macrophages, chondrocyte cell lines, primary chondrocytes, and synoviocytes, Heller et al. (1997; Proc Natl Acad Sci USA 94:2150-2155) identifed candidate genes involved in inflammatory responses, including TNF, IL-1 IL-6, IL-8 G-CSF, RANTES, and V-CAM. From this candidate gene set, tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase were found to be differentially expressed in rheumatoid arthritis (RA) relative to inflammatory bowel disease (IBD). Further, IL-3, chemokine Groα, and metalloproteinase matrix metallo-elastase were expressed in both RA and IBD. Most recently, in an analysis of cultured aortic smooth muscle cells treated with TNF-α, Haley et al. (2000; Circulation 102:2185-2189) found a 20-fold increase in eotaxin, an eosinophil chemotactic factor. The overexpression of eotaxin and its receptor CCR3 in atherosclerotic lesions was confirmed by northern analysis.

Human coronary artery endothelial cells (HCAECs) are primary cells derived from the endothelium of a human coronary artery. HCAECs are used as an experimental model for investigating the role of the endothelium in human vascular biology *in vitro*. Human umbilical artery endothelial cells (HUAECs) are primary cells derived from the endothelium of an umbilical artery. Human uterine myometrium microvascular endothelial cells (UtMVECs) are primary cells derived from the uterine myometrium microvasculature. Human Iliac Artery Endothelial Cells (HIAECs) are primary cells derived from the endothelium of an iliac artery. Human umbilical vein endothelial cells (HUVECs) are a primary cell line derived from the endothelium of the human umbilical vein. ECV304 is a human endothelial line.

Neurological disorders

Characterization of region-specific gene expression in the human brain provides a context and background for molecular neurobiology on a variety of neurological disorders. For example,

Alzheimer's disease (AD) is a progressive, neurodestructive process of the human neocortex, characterized by the deterioration of memory and higher cognitive function. A progressive and irreversible brain disorder, AD is characterized by three major pathogenic episodes involving (a) an aberrant processing and deposition of beta-amyloid precursor protein (betaAPP) to form neurotoxic beta-amyloid (betaA) peptides and an aggregated insoluble polymer of betaA that forms the senile plaque, (b) the establishment of intraneuronal neuritic tau pathology yielding widespread deposits of agyrophilic neurofibrillary tangles (NFT) and (c) the initiation and proliferation of a brain-specific inflammatory response. These three seemingly disperse attributes of AD etiopathogenesis are linked by the fact that proinflammatory microglia, reactive astrocytes and their associated cytokines and chemokines are associated with the biology of the microtubule associated protein tau, betaA speciation and aggregation. Missense mutations in the presenilin genes PS1 and PS2, implicated in early onset familial AD, cause abnormal betaAPP processing with resultant overproduction of betaA42 and related neurotoxic peptides. Specific betaA fragments such as betaA42 can further potentiate proinflammatory mechanisms. Expression of the inducible oxidoreductase cyclooxygenase-2 and cytosolic phospholipase A2 (cPLA2) is strongly activated during cerebral ischemia and trauma, epilepsy and AD, indicating the induction of proinflammatory gene pathways as a response to brain injury. Neurotoxic metals such as aluminum and zinc, both implicated in AD etiopathogenesis, and arachidonic acid, a major metabolite of brain cPLA2 activity, each polymerize hyperphosphorylated tau to form NFT-like bundles. Studies have identified a reduced risk for AD in patients aged over 70 years who were previously treated with non-steroidal anti-inflammatory drugs for non-CNS afflictions that include arthritis. (For a review of the interrelationships between the mechanisms of PS1, PS2 and betaAPP gene expression, tau and betaA deposition and the induction, regulation and proliferation in AD of the neuroinflammatory response, see Lukiw, W.J, and Bazan, N.G. (2000) Neurochem. Res. 2000 25:1173-1184).

Tumor necrosis factor-alpha (TNF-α) is a pleiotropic cytokine that plays a central role in mediation of the inflammatory response through activation of multiple signal transduction pathways. TNF-α is produced by activated lymphocytes, macrophages, and other white blood cells, and activates endothelial cells. Interferon-gamma (IFNγ), also known as Type II interferon or immune interferon, is a cytokine produced primarily by T-lymphocytes and natural killer cells. Mature IFNγ exists as noncovalently- linked homodimers. IFNγ displays antiviral, antiproliferative, immunoregulatory, and proinflammatory activities and is important in host defense mechanisms. IFN-γ induces the production of cytokines; upregulates the expression of class I and II MHC antigens, Fc receptor, and leukocyte adhesion molecule; modulates macrophage effector functions; influences isotype switching; potentiates

the secretion of immunoglobulins by B cells; augments TH1 cell expansion; and may be required for TH1 cell differentiation. IFNγ exerts its biological activities by binding to specific cell surface receptors, which display high affinity binding sites. The IFNγ receptor is present on almost all cell types except mature erythrocytes. Upon binding to its receptor, IFNγ triggers the activation of JAK-1 and JAK-2 kinases resulting in the phosphorylation of STAT1. Both IFNγ and TNF-α are considered proinflammatory cytokines. Cross-talk can exist between the signal transduction pathways of two cytokines; for example, signal transduction cascades initiated by two different cytokines lead to the activation of NfkB.

Liver toxicity

The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α-fetoprotein; iii) conversion of ammonia to urea and glutamine; iv) metabolism of aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. The C3A cell line is now well established as an *in vitro* model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am. J. Physiol. 272:G408-G416).

The potential application of gene expression profiling is relevant to measuring the toxic response to potential therapeutic compounds and of the metabolic response to therapeutic agents. For instance, diseases treated with steroids and disorders caused by the metabolic response to treatment with steroids include adenomatosis, cholestasis, cirrhosis, hemangioma, Henoch-Schonlein purpura, hepatitis, hepatocellular and metastatic carcinomas, idiopathic thrombocytopenic purpura, porphyria, sarcoidosis, and Wilson disease. It is desirable to measure the toxic response to potential therapeutic compounds and of the metabolic response to therapeutic agents.

Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and hormones, that share a common four-ring structure based on cyclopentanoperhydrophenanthrene and that carrry out a wide variety of functions. Steroid hormones, produced by the adrenal cortex, ovaries, and testes, include glucocorticoids, mineralocorticoids, androgens, and estrogens. Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for physical injuries and diseases such as arthritis, asthma, and auto-immune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive.

Medroxyprogesterone (MAH), also known as 6 α -methyl-17-hydroxyprogesterone, is a synthetic progestin with a pharmacological activity about 15 times greater than progesterone. MAH is usually used for the treatment of renal and endometrial carcinomas, amenorrhea, abnormal uterine bleeding, and endometriosis associated with hormonal imbalance. The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. The exact mechanism of action, however, is unknown. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, mammary gland, hypothalamus, and pituitary. Once bound to the receptor, progestins slow the frequency of release of gonadotropin releasing hormone (GnRH) from the hypothalamus and blunt the pre-ovulatory LH surge, thereby preventing follicular maturation and ovulation. Interestingly, the MAH stimulatory effect on the respiratory centers has been used clinically to treat low blood oxygenation due to sleep apnea, chronic obstructive pulmonary disease, or hypercapnia (excess of CO 2 in blood). Beclomethasone is a synthetic glucocorticoid that is used for treating steroid-dependent asthma, relieving symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or for preventing recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5,000 times greater than those produced by hydrocortisone. Budesonide is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. Prednisone is metabolized in the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties. Betamethasone is a synthetic glucocorticoid with antiinflammatory and immunosuppressive activity and is used to treat psoriasis and fungal infections, such as athlete's foot and ringworm. By comparing both the levels and sequences expressed in tissues from subjects exposed to or treated with steroid compounds with the levels and sequences expressed in normal untreated tissue it is possible to determine tissue responses to steroids. Budesonide (Bude) is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Budesonide has high topical anti-inflammatory activity but low systemic activity. Prednisone is a corticosteroid that is metabolized in the liver to its active form, prednisolone. Prednisone is roughly four times more potent as a glucocorticoid than hydrocortisone. Prednisone is intermediate between hydrocortisone and dexamethasone in duration of action. Prednisone is used in conditions such as allograft rejection, asthma, systemic lupus erythematosus, and many other inflammatory states.

Glucocorticoids are naturally occurring hormones that prevent or suppress inflammation and immune responses when administered at pharmacological doses. At the molecular level, unbound

glucocorticoids readily cross cell membranes and bind with high affinity to specific cytoplasmic receptors. Subsequent to binding, transcription and, ultimately, protein synthesis are affected. The result can include inhibition of leukocyte infiltration at the site of inflammation, interference in the function of mediators of inflammatory response, and suppression of humoral immune responses. The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A2 inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of transport, neurological, muscle, immunological and cell proliferative disorders.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, transporters and ion channels, referred to collectively as 'TRICH' and individually as 'TRICH-1,' 'TRICH-2,' 'TRICH-3,' 'TRICH-4,' 'TRICH-5,' 'TRICH-6,' 'TRICH-7,' 'TRICH-8,' 'TRICH-9,' 'TRICH-10,' 'TRICH-11,' 'TRICH-12,' 'TRICH-13,' 'TRICH-14,' 'TRICH-15,' 'TRICH-16,' 'TRICH-17,' 'TRICH-18,' 'TRICH-19,' 'TRICH-20,' 'TRICH-21,' 'TRICH-22,' 'TRICH-23,' 'TRICH-24,' 'TRICH-25,' and 'TRICH-26' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified transporters and ion channels and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified transporters and ion channels and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-26.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-26. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:27-52.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a

polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a

polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising

administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence

selected from the group consisting of SEQ ID NO:1-26. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing

the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a

host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TRICH" refers to the amino acid sequences of substantially purified TRICH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TRICH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

An "allelic variant" is an alternative form of the gene encoding TRICH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRICH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TRICH or a polypeptide with at least one functional characteristic of TRICH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRICH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding TRICH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino

acid residues which produce a silent change and result in a functionally equivalent TRICH. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TRICH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TRICH. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TRICH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to

immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a

naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic TRICH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding TRICH or fragments of TRICH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of TRICH or a polynucleotide encoding TRICH which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:27-52 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:27-52, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:27-52 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:27-52 from related polynucleotides. The precise length of a fragment of SEQ ID NO:27-52 and the region of SEQ ID NO:27-52 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-26 is encoded by a fragment of SEQ ID NO:27-52. A fragment of SEQ ID NO:1-26 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-26. For example, a fragment of SEQ ID NO:1-26 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-26. The precise length of a fragment of SEQ ID NO:1-26 and the region of SEQ ID NO:1-26 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g.,

methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version

2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by

CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in

the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune

disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TRICH which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TRICH which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TRICH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TRICH.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TRICH may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary

by cell type depending on the enzymatic milieu of TRICH.

"Probe" refers to nucleic acids encoding TRICH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989; Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to

avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TRICH, nucleic acids encoding TRICH, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral

infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between

individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides..

THE INVENTION

Various embodiments of the invention include new human transporters and ion channels (TRICH), the polynucleotides encoding TRICH, and the use of these compositions for the diagnosis, treatment, or prevention of transport, neurological, muscle, immunological and cell proliferative disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest

GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are transporters and ion channels. For example, SEQ ID NO:1 is 49% identical, from residue S11 to residue K626, to human CTL1 protein (GenBank ID g6996442) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.0e-168, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains an eight transmembrane helices regions as determined by using a hidden Markov model for the prediction of transmembrane helices. (See Table 3.) In an alternative example, SEQ ID NO:3 is 57% identical, from residue E10 to residue V115, to human SLC11A3 iron transporter (GenBank ID g8895485) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.7e-25, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. In an alternative example, SEQ ID NO:6 is 88% identical, from residue M1 to residue S944, to rat potassium channel (GenBank ID g2745729) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a PAC motif, a PAS domain, a cyclic nucleotide-binding domain, and an ion transport protein domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from

BLAST_PRODOM and BLAST_DOMO analyses provide further corroborative evidence that SEQ ID NO:6 is a potassium channel. In an alternative example, SEQ ID NO:10 is 99% identical, from residue M1 to residue I418, 95% identical, from residue S420 to residue S680, and 94% identical, from residue P665 to residue H894, to human Eag-related gene member 2 (GenBank ID g11878259) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 is localized to the plasma membrane, has transporter and channel activity and is a voltage-gated potassium channel, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:10 also contains a PAC domain, cyclic nucleotide binding domain and ion transport domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) SEQ ID NO:10 contains five transmembrane-spanning regions as determined by TMHMMR analysis. Data from further BLAST analyses of the PRODOM and DOMO databases provide additional corroborative evidence that SEQ ID NO:10 is a potassium channel. In an alternative example, SEQ ID NO:11 is 86% identical, from residue A94 to residue S785, to rat potassium channel (GenBank ID g2745729) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. As determined by BLAST analysis using the PROTEOME database, SEQ ID NO:11 is localized to the plasma membrane, is homologous to rat ether-a-go-go related 2, which is a slowly activating delayed rectifier potassium channel, and may facilitate the differentiation of pre-vertebral neurons (PROTEOME ID 331276|Rn.10875); SEQ ID NO:11 is also homologous to rat ether-a-go-go-related gene 3 which is an inward rectifier potassium channel that functions in potassium transport specifically in the nervous system (PROTEOME ID 331274|Rn.10874). SEQ ID NO:11 also contains a cyclic nucleotide-binding domain and an ion transport protein domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:11 is a potassium channel. In an alternative example, SEQ ID NO:14 is 38% identical, from residue Q13 to residue S1049, to Schizosaccharomyces pombe membrane ATPase (GenBank ID g3451312) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.5e-189, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:14 is

localized to the membrane, and is a member of the P-type, Ca2+-type, ATPase subfamily, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:14 also contains an E1-E2 ATPase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:14 is a membrane ATPase. In an alternative example, SEQ ID NO:19 is 98% identical, from residue M1 to residue L602, to human sodium-dependent high-affinity dicarboxylate transporter (GenBank ID g8132324) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:19 also has homology to proteins that have transporter gene function and are sodium-dependent dicarboxylate transporters, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:19 also contains a sodium-dependent dicarboxylate transporter domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and BLAST analyses provide further corroborative evidence that SEQ ID NO:19 is a sodium-dependent dicarboxylate transporter. SEQ ID NO:2, SEQ ID NO:4,-5, SEQ ID NO:7-9, SEQ ID NO:12-13, SEQ ID NO:15-18, and SEQ ID NO:20-26 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-26 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:8-14 or that distinguish between SEQ ID NO:8-14 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the

polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).

INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses TRICH variants. A preferred TRICH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TRICH amino acid sequence, and which contains at least one functional or structural characteristic of TRICH.

Various embodiments also encompass polynucleotides which encode TRICH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:27-52, which encodes TRICH. The polynucleotide

sequences of SEQ ID NO:27-52, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding TRICH. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding TRICH. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:27-52 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:27-52. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of TRICH.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding TRICH. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding TRICH, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding TRICH over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding TRICH. For example, a polynucleotide comprising a sequence of SEQ ID NO:34 and a polynucleotide comprising a sequence of SEQ ID NO:43 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:46 and a polynucleotide comprising a sequence of SEQ ID NO:52 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:39 and a polynucleotide comprising a sequence of SEQ ID NO:50 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:32, a polynucleotide comprising a sequence of SEQ ID NO:36, and a polynucleotide comprising a sequence of SEQ ID NO:37 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of TRICH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRICH, some bearing minimal

similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TRICH, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode TRICH and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring TRICH under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding TRICH or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRICH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode TRICH and TRICH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding TRICH or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:27-52 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines

such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding TRICH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode TRICH may be cloned in recombinant DNA molecules that direct expression of TRICH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express TRICH.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter TRICH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TRICH, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be

recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding TRICH may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, TRICH itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TRICH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active TRICH, the polynucleotides encoding TRICH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding TRICH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding TRICH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding TRICH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various

origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding TRICH and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding TRICH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding TRICH. For example, routine cloning, subcloning, and propagation of polynucleotides encoding TRICH can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding TRICH into the vector's multiple cloning site disrupts the *lacZ*

gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of TRICH are needed, e.g. for the production of antibodies, vectors which direct high level expression of TRICH may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TRICH. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of TRICH. Transcription of polynucleotides encoding TRICH may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding TRICH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses TRICH in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of

DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of TRICH in cell lines is preferred. For example, polynucleotides encoding TRICH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRICH is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding TRICH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRICH under the

control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding TRICH and that express TRICH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TRICH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TRICH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRICH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding TRICH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding TRICH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence

and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TRICH may be designed to contain signal sequences which direct secretion of TRICH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding TRICH may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TRICH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TRICH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TRICH encoding sequence and the heterologous protein sequence, so that TRICH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled TRICH may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple

transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

TRICH, fragments of TRICH, or variants of TRICH may be used to screen for compounds that specifically bind to TRICH. One or more test compounds may be screened for specific binding to TRICH. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to TRICH. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of TRICH can be used to screen for binding of test compounds, such as antibodies, to TRICH, a variant of TRICH, or a combination of TRICH and/or one or more variants TRICH. In an embodiment, a variant of TRICH can be used to screen for compounds that bind to a variant of TRICH, but not to TRICH having the exact sequence of a sequence of SEQ ID NO:1-26. TRICH variants used to perform such screening can have a range of about 50% to about 99% sequence identity to TRICH, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to TRICH can be closely related to the natural ligand of TRICH, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor TRICH (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to TRICH can be closely related to the natural receptor to which TRICH binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for TRICH which is capable of propagating a signal, or a decoy receptor for TRICH which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to TRICH, fragments of TRICH, or variants of TRICH. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of TRICH. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of TRICH. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of TRICH.

In an embodiment, anticalins can be screened for specific binding to TRICH, fragments of TRICH, or variants of TRICH. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit TRICH involves producing appropriate cells which express TRICH, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing TRICH or cell membrane fractions which contain TRICH are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TRICH or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TRICH, either in solution or affixed to a solid support, and detecting the binding of TRICH to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to

inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

TRICH, fragments of TRICH, or variants of TRICH may be used to screen for compounds that modulate the activity of TRICH. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TRICH activity, wherein TRICH is combined with at least one test compound, and the activity of TRICH in the presence of a test compound is compared with the activity of TRICH in the absence of the test compound. A change in the activity of TRICH in the presence of the test compound is indicative of a compound that modulates the activity of TRICH. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising TRICH under conditions suitable for TRICH activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TRICH may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TRICH or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and

the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding TRICH may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding TRICH can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TRICH is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TRICH, e.g., by secreting TRICH in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TRICH and transporters and ion channels. The expression of TRICH is closely associated with normal heart tissue, liver tumor tissue and diseased corpus callosum tissue. In addition, examples of tissues expressing TRICH can be found in Table 6 and can also be found in Example XI. Therefore, TRICH appears to play a role in transport, neurological, muscle, immunological and cell proliferative disorders. In the treatment of disorders associated with increased TRICH expression or activity, it is desirable to decrease the expression or activity of TRICH. In the treatment of disorders associated with decreased TRICH expression or activity, it is desirable to increase the expression or activity of TRICH.

Therefore, in one embodiment, TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis,

myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, glycogen storage disease, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, pseudohypoaldosteronism type 1, Liddle's syndrome, cystinuria, iminoglycinuria, Hartup disease, Fanconi disease, and Bartter syndrome; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, hemiplegic migraine, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's

muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, acid maltase deficiency (AMD, also known as Pompe's disease), generalized myotonia, and myotonia congenita; an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TRICH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent

a disorder associated with decreased expression or activity of TRICH including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TRICH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those listed above.

In a further embodiment, an antagonist of TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH. Examples of such disorders include, but are not limited to, those transport, neurological, muscle, immunological and cell proliferative disorders described above. In one aspect, an antibody which specifically binds TRICH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TRICH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TRICH may be produced using methods which are generally known in the art. In particular, purified TRICH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRICH. Antibodies to TRICH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels,

dromedaries, llamas, humans, and others may be immunized by injection with TRICH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TRICH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TRICH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRICH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRICH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for TRICH may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TRICH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRICH epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TRICH. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of TRICH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TRICH epitopes, represents the average affinity, or avidity, of the antibodies for TRICH. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular TRICH epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the TRICH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TRICH, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TRICH-antibody

complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding TRICH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TRICH. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRICH (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding TRICH may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated

cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in TRICH expression or regulation causes disease, the expression of TRICH from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TRICH are treated by constructing mammalian expression vectors encoding TRICH and introducing these vectors by mechanical means into TRICH-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TRICH include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TRICH may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TRICH from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental

parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TRICH expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TRICH under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding TRICH to cells which have one or more genetic abnormalities with respect to the expression of TRICH. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu.

Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TRICH to target cells which have one or more genetic abnormalities with respect to the expression of TRICH. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TRICH to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TRICH to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for TRICH into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TRICH-coding RNAs and the synthesis of high levels of TRICH in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy

application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TRICH into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding TRICH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding TRICH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA

polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TRICH. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TRICH expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TRICH may be therapeutically useful, and in the treatment of disorders associated with decreased TRICH expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TRICH may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TRICH is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TRICH are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is

detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TRICH. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TRICH, antibodies to TRICH, and mimetics, agonists, antagonists, or inhibitors of TRICH.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical,

sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TRICH or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TRICH or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRICH or fragments thereof, antibodies of TRICH, and agonists, antagonists or inhibitors of TRICH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are

used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind TRICH may be used for the diagnosis of disorders characterized by expression of TRICH, or in assays to monitor patients being treated with TRICH or agonists, antagonists, or inhibitors of TRICH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TRICH include methods which utilize the antibody and a label to detect TRICH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRICH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TRICH expression. Normal or standard values for TRICH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TRICH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TRICH expressed in

subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding TRICH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TRICH may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TRICH, and to monitor regulation of TRICH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding TRICH or closely related molecules may be used to identify nucleic acid sequences which encode TRICH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TRICH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TRICH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:27-52 or from genomic sequences including promoters, enhancers, and introns of the TRICH gene.

Means for producing specific hybridization probes for polynucleotides encoding TRICH include the cloning of polynucleotides encoding TRICH or TRICH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding TRICH may be used for the diagnosis of disorders associated with expression of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral

neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, glycogen storage disease, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, pseudohypoaldosteronism type 1, Liddle's syndrome, cystinuria, iminoglycinuria, Hartup disease, Fanconi disease, and Bartter syndrome; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, hemiplegic migraine, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear

myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, acid maltase deficiency (AMD, also known as Pompe's disease), generalized myotonia, and myotonia congenita; an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. Polynucleotides encoding TRICH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRICH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, polynucleotides encoding TRICH may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding TRICH may be labeled by standard methods and added to a fluid or tissue

sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding TRICH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRICH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRICH, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRICH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding TRICH, or a fragment of a polynucleotide complementary to the polynucleotide encoding TRICH, and will be employed under optimized conditions for identification of a specific gene or

condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding TRICH may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding TRICH are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations

and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of TRICH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, TRICH, fragments of TRICH, or antibodies specific for TRICH may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present

invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the

proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TRICH to quantify the levels of TRICH expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which

alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding TRICH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-

355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TRICH on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TRICH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRICH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TRICH, or fragments thereof, and washed. Bound TRICH is then detected by methods well known in the art. Purified TRICH can also be

coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRICH specifically compete with a test compound for binding TRICH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TRICH.

In additional embodiments, the nucleotide sequences which encode TRICH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/313,242, U.S. Ser. No. 60/324,782, U.S. Ser. No. 60/328,184, U.S. Ser. No. 60/345,937, U.S. Ser. No. 60/335,698, U.S. Ser. No. 60/332,804, U.S. Ser. No. 60/333,922, U.S. Ser. No. 60/388,180, U.S. Ser. No. 60/375,637, and U.S. Ser. No. 60/377,444, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and

ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a

high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus

primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:27-52. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative transporters and ion channels were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transporters and ion channels, the encoded polypeptides were analyzed by querying against PFAM models for transporters and ion channels. Potential transporters and ion channels were also identified by homology to Incyte cDNA sequences that had been annotated as transporters and ion channels. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to

be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of TRICH Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:27-52 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:27-52 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences

had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by

gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding TRICH are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TRICH. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of TRICH Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68 °C to about 72 °C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one

extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and

4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in TRICH Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:27-52 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown

of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:27-52 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470;

Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte Genomics). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are

amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-

scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte Genomics). Array elements that exhibited at least about a two-fold change in expression, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed.

Expression

For example, SEQ ID NO:30 showed differential expression in bone osteosarcoma tissues

versus normal osteocytes, as determined by microarray analysis. Messenger RNA from normal human osteoblasts was compared with mRNA from biopsy specimens, osteosarcoma tissues, primary cultures, or metastasized tissues. A normal osteoblast primary culture, NHOst 5488, was chosen as the reference in the initial experiments. One basic set of experiments is defined as the comparison of mRNA from biopsy specimen with that of definitive surgical specimen from the same patient. Extended study of this basic set includes mRNA from primary cell cultures of the definitive surgical specimen, muscle, or cartilage tissue from the same patient. Biopsy specimens, definitive surgical specimens, or lung metastatic tissues from different individuals were also included to reveal individual variability. The expression of SEQ ID NO:30 was increased by at least two-fold in bone osteosarcoma tissues relative to normal osteocytes. Therefore, in various embodiments, SEQ ID NO:30 can be used for one or more of the following: i) monitoring treatment of bone cancer, ii) diagnostic assays for bone cancer, and iii) developing therapeutics and/or other treatments for bone cancer.

In an alternative example, SEQ ID NO:33 showed differential expression in lung squamous carcinoma tissues versus normal lung tissues as determined by microarray analysis. In matched tissue experiments, the expression of SEQ ID NO:33 was decreased by at least two-fold in lung squamous carcinoma tissues relative to grossly uninvolved normal lung tissues from the same donors. Therefore, in various embodiments, SEQ ID NO:33 can be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

SEQ ID NO:33 also showed differential expression in ovarian adenocarcinoma tissues versus normal ovarian tissues as determined by microarray analysis. The expression of SEQ ID NO:33 was decreased by at least two-fold in ovarian adenocarcinoma tissues relative to grossly uninvolved normal ovarian tissues from the same donor. Therefore, in various embodiments, SEQ ID NO:33 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

SEQ ID NO:33 and SEQ ID NO:40 showed differential expression in association with immune and inflammatory responses as determined by microarray analysis. The expression of SEQ ID NO:33 was increased by at least two-fold in human umbilical vein cells treated with PMA and ionomycin relative to untreated human umbilical vein cells. Human umbilical vein cells are derived from the endothelium of the human umbilical vein, and have been used as an experimental model for investigating the functional biology of human endothelial cells in vitro. PMA is a broad activator of

protein kinase C-dependent pathways and ionomycin is a calcium ionophore that permits the entry of calcium in the cell, hence increasing the cytosolic calcium concentration. The expression of SEQ ID NO:40 was increased by at least 2.5-fold in vascular endothelial tissue treated with TNFα and IL-1β compared with untreated vascular endothelial tissue, as determined by microarray analysis. Human coronary artery endothelial cells and human coronary artery smooth muscle cells (BioWhittaker, Inc., San Diego CA) obtained from the same donor were cultured in tissue culture flasks in Endothelium Growth Medium or Smooth Muscle Growth Medium, respectively (BioWhittaker). Cultures at 85% confluency were either treated with recombinant human TNFα and IL-1β (R&D Systems, Minneapolis MN) at 10 ng/ml each for 24 hours at 37° C or were left untreated. Therefore, in various embodiments, SEQ ID NO:33 and SEQ ID NO:40 can each be used for one or more of the following: i) monitoring treatment of immune/inflammatory responses, ii) diagnostic assays for immune/inflammatory responses, and iii) developing therapeutics and/or other treatments for immune/inflammatory responses.

In an alternative example, SEQ ID NO:38 showed differential expression in breast carcinoma cell lines versus primary mammary epithelial cells as determined by microarray analysis. The breast carcinoma cell lines include MCF7, a breast adenocarcinoma cell line derived from the pleural effusion of a 69-year-old female; T-47D, a breast carcinoma cell line derived from a pleural effusion from a 54-year-old female with an infiltrating ductal carcinoma of the breast; Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female; MDA-mb-231, a metastatic breast tumor cell line derived from the pleural effusion of a 51-year-old female with metastatic breast carcinoma; and MDA-mb-435S, a spindle shaped strain that evolved from a cell line isolated from the pleural effusion of a 31 year old female with metastatic, ductal adenocarcinoma of the breast. The primary mammary epithelial cell line HMEC was derived from normal human mammary tissue (Clonetics, San Diego, CA). All cell cultures were propagated in a chemically-defined medium, according to the supplier's recommendations and grown to 70-80% confluence prior to RNA isolation. The microarray experiments showed that the expression of SEQ ID NO:38 was decreased by at least two-fold in all five breast carcinoma lines (MCF7, T-47D, Sk-BR-3, MDA-mb-231, and MDA-mb-435S) relative to primary mammary epithelial cells. Therefore, in various embodiments, SEQ ID NO:38 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

SEQ ID NO:38 also showed differential expression in certain prostate carcinoma cell lines

versus normal prostate epithelial cells as determined by microarray analysis. The prostate carcinoma cell lines include CA-HPV-10, DU 145, LNCaP, PC-3, and MDAPCa2b. CA-HPV-7 was derived from cells from a 63 year old male with prostate adenocarcinoma and was transformed by transfection with HPV18 DNA. DU 145 was isolated from a metastatic site in the brain of a 69 year old male with widespread metastatic prostate carcinoma. DU 145 has no detectable sensitivity to hormones; forms colonies in semi-solid medium; is only weakly positive for acid phosphatase; and cells are negative for prostate specific antigen (PSA). LNCaP is a prostate carcinoma cell line isolated from a lymph node biopsy of a 50 year old male with metastatic prostate carcinoma. LNCaP expresses PSA, produces prostate acid phosphatase, and expresses androgen receptors. PC-3, a prostate adenocarcinoma cell line, was isolated from a metastatic site in the bone of a 62 year old male with grade IV prostate adenocarcinoma. MDAPCa2b is a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 63 year old male. The MDAPCa2b cell line expresses PSA and androgen receptor and is androgen sensitive. The normal epithelial cell line, PrEC, is a primary prostate epithelial cell line isolated from a normal donor. The expression of SEQ ID NO:38 was decreased by at least two-fold in three out of five prostate carcinoma lines (DU 145, LNCaP, and PC-3) relative to cells from the normal prostate epithelial cell line, PrEC. Therefore, in various embodiments, SEQ ID NO:38 can be used for one or more of the following: i) monitoring treatment of prostate cancer, ii) diagnostic assays for prostate cancer, and iii) developing therapeutics and/or other treatments for prostate cancer.

In addition, SEQ ID NO:38 and SEQ ID NO:43 showed differential expression in toxicology studies as determined by microarray analysis. The expression of SEQ ID NO:43 was increased by at least two-fold in C3A hepatoblastoma cells treated with 1-100 μM beclomethazone as compared with untreated C3A hepatoblastoma cells. The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α-fetoprotein iii) conversion of ammonia to urea and glutamine; iv) metabolism of aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. The C3A cell line is now well established as an in vitro model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am J Physiol 272:G408-G416). C3A cells were treated with 1-100 μM beclomethazone for 1hr, 3hr, 6hr and compared with untreated cells. In

addition, the expression of SEQ ID NO:34 was increased by at least two-fold in early confluent C3A cells treated with progesterone, beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, or betamethasone, for 1, 3, or 6 hours, as compared to untreated C3A cells. In addition, the expression of SEQ ID NO:38 was decreased by at least two-fold in early confluent C3A cells treated with progesterone, beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, or betamethasone, for 1, 3, or 6 hours, as compared to untreated C3A cells. The effects of steroids on liver metabolism are important to the understanding of the pharmacodynamics of drugs. Therefore, in various embodiments, SEQ ID NO:34, SEQ ID NO:38 and SEQ ID NO:43 can each be used for one or more of the following: i) monitoring treatment of liver toxicity, diseases and disorders, ii) diagnostic assays for liver toxicity, diseases and disorders, and iii) developing therapeutics and/or other treatments for liver toxicity, diseases and disorders.

In yet another example, the expression of SEQ ID NO:48 was differentially expressed in a specific region of human brain tissue as compared to pooled brain tissue control. Characterization of region-specific gene expression in the human brain provides a context and background for molecular neurobiology research in general. This knowledge may provide insight into the genetic basis of brain structure and function. The expression of SEQ ID NO:48 was decreased by at least two-fold in normal human amygdala, entorhinal cortex, brain tissue as compared to the normal human pooled brain tissue used as a control. These experiments indicate that SEQ ID NO:48 exhibited significant differential expression patterns using microarray techniques, and further establishes its utility as a diagnostic marker or therapeutic agent which may be useful in neurological disorders. Therefore, in various embodiments, SEQ ID NO:48 can be used for one or more of the following: i) monitoring treatment of neurological disorders, ii) diagnostic assays for neurological disorders, and iii) developing therapeutics and/or other treatments for neurological disorders.

XII. Complementary Polynucleotides

Sequences complementary to the TRICH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRICH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TRICH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRICH-encoding transcript.

XIII. Expression of TRICH

Expression and purification of TRICH is achieved using bacterial or virus-based expression systems. For expression of TRICH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TRICH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TRICH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TRICH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, TRICH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from TRICH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified TRICH obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

TRICH function is assessed by expressing the sequences encoding TRICH at physiologically

elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of TRICH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TRICH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TRICH and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of TRICH Specific Antibodies

TRICH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the TRICH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TRICH activity by, for example, binding the peptide or TRICH to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring TRICH Using Specific Antibodies

Naturally occurring or recombinant TRICH is substantially purified by immunoaffinity chromatography using antibodies specific for TRICH. An immunoaffinity column is constructed by covalently coupling anti-TRICH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TRICH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TRICH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRICH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRICH is collected.

XVII. Identification of Molecules Which Interact with TRICH

TRICH, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, molecules interacting with TRICH are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially

available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Identification of Molecules Which Interact with TRICH

Molecules which interact with TRICH may include transporter substrates, agonists or antagonists, modulatory proteins such as Gβγ proteins (Reimann, *supra*) or proteins involved in TRICH localization or clustering such as MAGUKs (Craven, *supra*). TRICH, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, proteins that interact with TRICH are isolated using the yeast 2-hybrid system (Fields, S. and O. Song (1989) Nature 340:245-246). TRICH, or fragments thereof, are expressed as fusion proteins with the DNA binding domain of Gal4 or lexA, and potential interacting proteins are expressed as fusion proteins with an activation domain. Interactions between the TRICH fusion protein and the TRICH interacting proteins (fusion proteins with an activation domain) reconstitute a transactivation function that is observed by expression of a reporter gene. Yeast 2-hybrid systems are commercially available, and methods for use of the yeast 2-hybrid system with ion channel proteins are discussed in Niethammer, M. and M. Sheng (1998, Meth. Enzymol. 293:104-122).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential TRICH agonists or antagonists may be tested for activation or inhibition of TRICH ion channel activity using the assays described in section XVIII.

XVIII. Demonstration of TRICH Activity

Ion channel activity of TRICH is demonstrated using an electrophysiological assay for ion conductance. TRICH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TRICH. Eukaryotic expression vectors are

commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A second plasmid which expresses any one of a number of marker genes, such as ß-galactosidase, is co-transformed into the cells to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of TRICH and ß-galactosidase.

Transformed cells expressing ß-galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or ß-galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TRICH will have higher anion or cation conductance relative to control cells. The contribution of TRICH to conductance can be confirmed by incubating the cells using antibodies specific for TRICH. The antibodies will bind to the extracellular side of TRICH, thereby blocking the pore in the ion channel, and the associated conductance.

Alternatively, ion channel activity of TRICH is measured as current flow across a TRICH-containing *Xenopus laevis* oocyte membrane using the two-electrode voltage-clamp technique (Ishi et al., *supra*; Jegla, T. and L. Salkoff (1997) J. Neurosci. 17:32-44). TRICH is subcloned into an appropriate *Xenopus* oocyte expression vector, such as pBF, and 0.5-5 ng of mRNA is injected into mature stage IV oocytes. Injected oocytes are incubated at 18°C for 1-5 days. Inside-out macropatches are excised into an intracellular solution containing 116 mM K-gluconate, 4 mM KCl, and 10 mM Hepes (pH 7.2). The intracellular solution is supplemented with varying concentrations of the TRICH mediator, such as cAMP, cGMP, or Ca^{+2} (in the form of $CaCl_2$), where appropriate. Electrode resistance is set at 2-5 MΩ and electrodes are filled with the intracellular solution lacking mediator. Experiments are performed at room temperature from a holding potential of 0 mV. Voltage ramps (2.5 s) from -100 to 100 mV are acquired at a sampling frequency of 500 Hz. Current measured is proportional to the activity of TRICH in the assay.

Transport activity of TRICH is assayed by measuring uptake of labeled substrates into *Xenopus laevis* oocytes. Oocytes at stages V and VI are injected with TRICH mRNA (10 ng per oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 1mM Na_2HPO_4, 5 mM Hepes, 3.8 mM NaOH , 50μg/ml gentamycin, pH 7.8) to allow expression of TRICH. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2

mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, ions, and neurotransmitters) is initiated by adding labeled substrate (e.g. radiolabeled with ^3H, fluorescently labeled with rhodamine, etc.) to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+-free medium, measuring the incorporated label, and comparing with controls. TRICH activity is proportional to the level of internalized labeled substrate.

ATPase activity associated with TRICH can be measured by hydrolysis of radiolabeled ATP-[γ-^{32}P], separation of the hydrolysis products by chromatographic methods, and quantitation of the recovered ^{32}P using a scintillation counter. The reaction mixture contains ATP-[γ-^{32}P] and varying amounts of TRICH in a suitable buffer incubated at 37°C for a suitable period of time. The reaction is terminated by acid precipitation with trichloroacetic acid and then neutralized with base, and an aliquot of the reaction mixture is subjected to membrane or filter paper-based chromatography to separate the reaction products. The amount of ^{32}P liberated is counted in a scintillation counter. The amount of radioactivity recovered is proportional to the ATPase activity of TRICH in the assay.

Lipocalin activity of TRICH is measured by ligand fluorescence enhancement spectrofluorometry (Lin et al. (1997) Molecular Vision 3:17). Examples of ligands include retinol (Sigma, St. Louis MO) and 16-anthryloxy-palmitic acid (16-AP) (Molecular Probes Inc., Eugene OR). Ligand is dissolved in 100% ethanol and its concentration is estimated using known extinction coefficents (retinol: 46,000 A/M/cm at 325 nm; 16-AP: 8,200 A/M/cm at 361 nm). A 700 μl aliquot of 1 μM TRICH in 10 mM Tris (pH 7.5), 2 mM EDTA, and 500 mM NaCl is placed in a 1 cm path length quartz cuvette and 1 μl aliquots of ligand solution are added. Fluorescence is measured 100 seconds after each addition until readings are stable. Change in fluorescence per unit change in ligand concentration is proportional to TRICH activity.

XIX. Identification of TRICH Agonists and Antagonists

TRICH is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293. Ion channel activity of the transformed cells is measured in the presence and absence of candidate agonists or antagonists. Ion channel activity is assayed using patch clamp methods well known in the art or as described in Example XVIII. Alternatively, ion channel activity is assayed using fluorescent techniques that measure ion flux across the cell membrane (Velicelebi, G. et al. (1999) Meth. Enzymol. 294:20-47; West, M.R. and C.R. Molloy (1996) Anal. Biochem. 241:51-58). These assays may be adapted for high-throughput screening using microplates. Changes in internal ion concentration are measured using fluorescent dyes such as the

Ca^{2+} indicator Fluo-4 AM, sodium-sensitive dyes such as SBFI and sodium green, or the Cl^- indicator MQAE (all available from Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In a more generic version of this assay, changes in membrane potential caused by ionic flux across the plasma membrane are measured using oxonyl dyes such as $DiBAC_4$ (Molecular Probes). $DiBAC_4$ equilibrates between the extracellular solution and cellular sites according to the cellular membrane potential. The dye's fluorescence intensity is 20-fold greater when bound to hydrophobic intracellular sites, allowing detection of $DiBAC_4$ entry into the cell (Gonzalez, J.E. and P.A. Negulescu (1998) Curr. Opin. Biotechnol. 9:624-631). Candidate agonists or antagonists may be selected from known ion channel agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
1853191	1	1853191CD1	27	1853191CB1	
7497369	2	7497369CD1	28	7497369CB1	3508438CA2
1700438	3	1700438CD1	29	1700438CB1	1700438CA2, 90166551CA2, 90167174CA2, 90167190CA2
535939	4	535939CD1	30	535939CB1	90129304CA2
55118067	5	55118067CD1	31	55118067CB1	
7502087	6	7502087CD1	32	7502087CB1	
7500819	7	7500819CD1	33	7500819CB1	
7503413	8	7503413CD1	34	7503413CB1	
7500007	9	7500007CD1	35	7500007CB1	
7500025	10	7500025CD1	36	7500025CB1	
7502736	11	7502736CD1	37	7502736CB1	
7503570	12	7503570CD1	38	7503570CB1	
7504008	13	7504008CD1	39	7504008CB1	90010166CA2, 90010266CA2, 90010442CA2
7503559	14	7503559CD1	40	7503559CB1	90198141CA2
6243872	15	6243872CD1	41	6243872CB1	
90011608	16	90011608CD1	42	90011608CB1	90011075CA2, 90011083CA2, 90011608CA2, 90011624CA2
90024583	17	90024583CD1	43	90024583CB1	90024583CA2
90113658	18	90113658CD1	44	90113658CB1	
3942766	19	3942766CD1	45	3942766CB1	

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7501987	20	7501987CD1	46	7501987CB1	5842557CA2, 90119749CA2, 90119849CA2, 90119857CA2, 90120041CA2, 90152509CA2
7503223	21	7503223CD1	47	7503223CB1	90041459CA2, 90041551CA2
7503566	22	7503566CD1	48	7503566CB1	90066119CA2, 90066211CA2
7505122	23	7505122CD1	49	7505122CB1	1803394CA2, 1805561CA2
7511620	24	7511620CD1	50	7511620CB1	
7506995	25	7506995CD1	51	7506995CB1	90119757CA2
7506996	26	7506996CD1	52	7506996CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	1853191CD1	g6996442	9.0E-168	[Homo sapiens] CTL1 protein
				O'Regan, S. et al., An electric lobe suppressor for a yeast choline transport mutation belongs to a new family of transporter-like proteins. Proc. Natl. Acad. Sci. U.S.A. 97 (4), 1835-1840 (2000).
2	7497369CD1	g9663117	6.7E-44	[Homo sapiens] organic cation transporter
3	1700438CD1	g8895485	4.7E-25	[Homo sapiens] SLC11A3 iron transporter
				Abboud, S. et al., A novel mammalian iron-regulated protein involved in intracellular iron metabolism J. Biol. Chem. 275 (26), 19906-19912 (2000).
4	535939CD1	g2621849	2.9E-47	[Methanothermobacter thermautotrophicus] Na+/H+-exchanging protein:Na+/H+ antiporter
				Smith, D.R., et al., Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics J. Bacteriol. 179 (22), 7135-7155 (1997).
5	55118067CD1	g9758059	5.0E-25	[Arabidopsis thaliana] amino acid transporter-like protein
				Sato, S. et al., Structural analysis of Arabidopsis thaliana chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and TAC clones. DNA Res. 5 (1), 41-54 (1998).
6	7502087CD1	g2745729	0.0	[Rattus norvegicus] potassium channel
				Shi, W. et al., Identification of two nervous system-specific members of the erg potassium channel gene family. J. Neurosci. 17 (24), 9423-9432 (1997).
7	7500819CD1	g17391438	9.0E-31	[Homo sapiens] FXYD domain-containing ion transport regulator 6
		g15216228	2.9E-24	[Mus musculus] (AB032010) PLM-like protein
8	7503413CD1	g3065814	2.3E-160	[Rattus norvegicus] sodium-dicarboxylate cotransporter SDCT1
				Chen, X.Z. et al., Characterization of a rat Na+-dicarboxylate cotransporter. J. Biol. Chem. 273 (33), 20972-20981 (1998).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
8 cont.		338046| SLC13A2	8.6E-161	[Homo sapiens][Active transporter, secondary; Transporter] [Plasma membrane] Rena Na(+)-dicarboxylate cotransporter, reabsorbs Krebs cycle intermediates, such as succinate and citrate, from the glomerular filtrate.
9	7500007CD1	g202523	1.4E-261	[Rattus norvegicus] GABA transporter Borden,L.A., et al. (1992) J. Biol. Chem. 267:21098-21104. Molecular heterogeneity of the GABA transport System: Cloning of two novel high-affinity GABA transporters from rat brain.
		570794| SLC6A13	2.5E-220	[Homo sapiens] Gamma-aminobutyric acid (GABA) transporter, expressed in brain, retina, and peripheral tissues, requires sodium and chloride for activity.
10	7500025CD1	g11878259	0.0	[Homo sapiens] Eag-related gene member 2
		571102| KCNH2	1.3E-290	[Homo sapiens][Channel (passive transporter); Transporter; Small molecule-binding protein][Plasma membrane] Voltage-gated (delayed rectifier) potassium channel, forms the Ikr channels that are important for cardiac rhythm; mutations in the gene are associated with the long QT form of cardiac arrhythmia.
11	7502736CD1	g2745729	0.0	[Rattus norvegicus] potassium channel (Shi, W. et al. (1997) Identification of two nervous system-specific members of the erg potassium channel gene family. J. Neurosci. 17:9423-9432.)
		331276| Rn.10875	0.0	[Rattus norvegicus][Channel (passive transporter); Transporter][Plasma membrane] Ether-a-go-go related 2, a member of the erg family of potassium channels, contains six putative transmembrane domains, a slowly activating delayed rectifier potassium channel, may facilitate the differentiation of pre-vertebral neurons (Shi, W. et al. (1997) supra.)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
11 cont.		331274| Rn.10874	5.2E-233	[Rattus norvegicus][Channel (passive transporter);Transporter][Plasma membrane] Ether-a-go-go-related gene 3, an inward rectifier potassium channel that functions in potassium transport specifically in the nervous system (Shi, W. et al. (1997) supra.)
12	7503570CD1	g5926734	9.6E-265	[Homo sapiens] 4F2 heavy chain
		567834| SLC3A2	1.5E-259	[Homo sapiens][Active transporter, secondary;Transporter][Plasma membrane] Heavy chain of 4F2 cell-surface antigen, a component of system L amino acid transport activity, involved in T-cell activation, cell proliferation and the regulation of HIV-mediated cell fusion; over-expression is associated with malignant transformation (Rajan, D. P. et al. (2000) Differential influence of the 4F2 heavy chain and the protein related to b(0,+) amino acid transport on substrate affinity of the heteromeric b(0,+) amino acid transporter. J. Biol. Chem. 275:14331-14335.)
		321548| Mdu1	4.0E-200	[Mus musculus][Protein kinase; Transferase; Active transporter, secondary; Transporter; Receptor (signaling)][Plasma membrane] Heavy chain of 4F2 cell-surface antigen, a component of system L amino acid transport activity, involved in cellular activation and proliferation; over-expression of human MDU1 (SLC3A2) is associated with malignant transformation (Rajan, D. P. et al. (2000) supra.)
13	7504008CD1	g13344999	0.0	[Homo sapiens] solute carrier family 26 member 6 (Waldegger, S. et al. (2001) Cloning and Characterization of SLC26A6, a Novel Member of the Solute Carrier 26 Gene Family. Genomics 72:43-50.)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
13 cont.		662428| SLC26A6	0.0	[Homo sapiens] Member of a family of anion transporters that exchange chloride and bicarbonate ions across membranes in an electroneutral manner, expressed on apical plasma membranes of duct cells in pancreas. (Lohi, H. et al. (2000) Mapping of five new putative anion transporter genes in human and characterization of SLC26A6, A candidate gene for pancreatic anion exchanger. Genomics 70, 102-112; Waldegger, S. et al. (2001) supra.).
		336854| SLC26A4	2.8E-121	[Homo sapiens][Transporter] Pendrin, a chloride and iodide transporter; mutation of corresponding gene causes Pendred syndrome and a common form of non-syndromic recessive deafness. (Scott, D. A. et al. (1999) The Pendred syndrome gene encodes a chloride-iodide transport protein. Nat. Genet. 21:440-443; Usami, S. et al. (1999) Non-syndromic hearing loss associated with enlarged vestibular aqueduct is caused by PDS mutations. Hum. Genet. 104:188-192.)
		339298| SLC26A3	9.0E-112	[Homo sapiens][Transcription factor; Transporter] Solute carrier family 26 member 3, involved in anion and sulfate transport, down-regulated in colon adenomas and adenocarcinoma; gene mutation causes congenital chloride diarrhea. (Hoglund, P. et al. (1998) Clustering of private mutations in the congenital chloride diarrhea/down-regulated in adenoma gene. Hum. Mutat. 11:321-327.)
14	7503559CD1	g3880560	5.6E-199	[Caenorhabditis elegans] Similarity to Yeast E1-E2 ATPase (SW:YED1_YEAST), contains similarity to Pfam domain: PF00122 (E1-E2 ATPase), Score=102.4, E-value=2.7e-28
		313777| K07E3.7	7E-204	[Caenorhabditis elegans] [Hydrolase; ATPase] Member of the P-type ATPase, Ca2+-type subfamily protein family.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
14 cont.		251962| W08D2.5	4.9E-200	[Caenorhabditis elegans] [Active transporter, primary; Hydrolase; Transporter; ATPase] [Unspecified membrane] Member of the P-type ATPase, Ca2+-type subfamily protein family.
		9082| YOR291W	5.2E-192	[Saccharomyces cerevisiae] [Active transporter, primary; Hydrolase; Transporter; ATPase] [Unspecified membrane] Member of the cation-translocating P-type ATPase superfamily of membrane transporters.
15	6243872CD1	g14485581	2E-150	[Rattus norvegicus] testis-specific transporter TST1
		617752| SLC21A12	1.7E-85	[Homo sapiens] [Transporter] [Unspecified membrane] Organic anion transporter (solute carrier family 21 member 12), a member ofthe OATP family that transports thyroid hormone by a Na+independent process, may be involved in the transport of bile acids and prostaglandins. Fujiwara, K. et al. (2001) Endocrinology 142:2005-2012 Identification of thyroid hormone transporters in humans: different molecules are involved in a tissue-specific manner.
		252769| Y70G10A.3	1.3E-58	[Caenorhabditis elegans] [Transporter] [Unspecified membrane] Member of the prostaglandin transporter protein family.
		343800| SLC21A6	2E-52	[Homo sapiens] [Active transporter, secondary; Transporter] [Unspecified membrane; Plasma membrane] Organic anion transporter, has a broad substrate specificity and is expressed in the liver.
		618842| Slc21a14	4.2E-52	[Mus musculus] [Transporter] Protein with high similarity to human SLC21A6, which is an organic anion transporter with broad substrate specificity and expressed in the liver.
		245952| F53B1.8	5.8E-50	[Caenorhabditis elegans] [Transporter] [Unspecified membrane] Member of the prostaglandin transporter protein family.
16	90011608CD1	g14571904	8.8E-222	[Rattus norvegicus] lysosomal amino acid transporter 1 Sagne, C. et al. (2001) Proc. Natl. Acad. Sci. U S A 98:7206-7211 Identification and characterization of a lysosomal transporter for small neutral amino acids.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
16 cont.		252408| Y43F4B.7	2.2E-65	[Caenorhabditis elegans] [Active transporter, secondary; Transporter] [Unspecified membrane] Putative membrane transporter of the amino acid/auxin permease-like (AAAP-like) protein family.
		251329| T27A1.5	3E-63	[Caenorhabditis elegans] Protein with similarity to membrane transporters of the amino acid/auxin permease family, putative paralog of C. elegans Y43F4B.7.
		240958| C44B7.6	3.1E-43	[Caenorhabditis elegans] [Active transporter, secondary; Transporter] [Unspecified membrane] Member of the amino acid/auxin permease protein family.
		247128| H32K16.1	2.8E-42	[Caenorhabditis elegans] Putative paralog of C.elegans Y38H6C.17, has weak similarity to C. elegans UNC-47, vesicular GABA transporter (VGAT).
		313559| F59B2.2	4.8E-40	[Caenorhabditis elegans] Protein with strong similarity to C. elegans Y38H6C.17, has similarity to C. elegans UNC-47, vesicular GABA transporter.
17	90024583CD1	g1098557	4.8E-166	[Homo sapiens] renal sodium/dicarboxylate cotransporter
		338046| SLC13A2	4.2E-167	[Homo sapiens] [Active transporter, secondary; Transporter] [Unspecified membrane; Plasma membrane] Sodium-dependent dicarboxylate cotransporter (solute carrier family 13 member 2), reabsorbs Krebs cycle intermediates, such as succinate and citrate, from the renal tubular filtrate. Pajor, A.M. (1996) Am. J. Physiol. 270:F642-648 Molecular cloning and functional expression of a sodium-dicarboxylate cotransporter from human kidney.
		331190| Rn.10821	1E-165	[Rattus norvegicus] [Active transporter, secondary; Transporter] [Unspecified membrane] Sodium-dependent dicarboxylate cotransporter (solute carrier family 13 member 2), reabsorbs Krebs cycle intermediates, such as succinate and citrate, from the renal tubular filtrate.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
17 cont.		742868| SLC13A3	1.3E-142	[Homo sapiens] Na(+)-dependent high-affinity dicarboxylate transporter 3, mediates succinate transport in the presence of sodium.
		661462| Nadc3	9.2E-140	[Rattus norvegicus] [Active transporter, secondary; Transporter] [Unspecified membrane] Sodium-dependent high-affinity dicarboxylate transporter 3, may function in di- andtri-carboxylic acid transport and elimination of organic anions and therapeutic drugs in the kidney.
		742872| SLC13A4	2.3E-132	[Homo sapiens] Sulfate transporter, mediates sodium-dependent sulfate transport at high levels, may function in sulfate incorporation in high endothelial venules.
18	90113658CD1	g6010763	8.1E-180	[Rattus norvegicus] ion transporter protein
		430266| Slc22a3	5.8E-29	[Mus musculus] [Active transporter, secondary; Major Facilitator Superfamily; Transporter] [Unspecified membrane; Plasma membrane] Solute carrier family 22 member 3 (extraneuronal monoamine transporter), regulates monoamine transport in the heart and placenta. Grundemann, D. et al. (1999) Selective substrates for non-neuronal monoamine transporters. Mol. Pharmacol. 56:1-10.
19	3942766CD1	g8132324	0.0	[Homo sapiens] sodium-dependent high-affinity dicarboxylate transporter. Wang, H. et al. (2000) Structure, function, and genomic organization of human Na+-dependent high-affinity dicarboxylate transporter. Am. J. Physiol. 278:C1019-C1030.
		742868| SLC13A3	0.0	[Homo sapiens] Na(+)-dependent high-affinity dicarboxylate transporter 3, mediates succinate transport in the presence of sodium.
		661462| Nadc3	2.6E-286	[Rattus norvegicus] [Active transporter, secondary; Transporter] [Unspecified membrane] Sodium-dependent high-affinity dicarboxylate transporter 3, may function in di- and tri-carboxylic acid transport and elimination of organic anions and therapeutic drugs in the kidney. (Wang, H. et al. (2000) supra.)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
19 cont.		331190| Rn.10821	3.2E-137	[Rattus norvegicus] [Active transporter,secondary; Transporter] [Unspecified membrane] Sodium-dependent dicarboxylate co-transporter (solute carrier family 13 member 2), reabsorbs Krebs cycle intermediates, such as succinate and citrate, from the renal tubular filtrate.
		338046| SLC13A2	8.8E-135	[Homo sapiens] [Active transporter, secondary; Transporter] [Unspecified membrane; Plasma membrane] Sodium-dependent dicarboxylate cotransporter (solute carrier family 13 member 2), reabsorbs Krebs cycle intermediates, such as succinate and citrate, from the renal tubular filtrate.
		742872| SLC13A4	1.7E-121	[Homo sapiens] Sulfate transporter, mediates sodium-dependent sulfate transport at high levels, may function insulfate incorporation in high endothelial venules.
20	7501987CD1	g20384654	1.0E-151	[Homo sapiens] two-pore calcium channel protein 2
		g14041819	8.5E-17	[Arabidopsis thaliana] (AB053952) two-pore calcium channel. Furuichi,T. et al. (2001) A Putative Two Pore Channel AtTPC1 Mediates Ca(2+) Flux in Arabidopsis Leaf Cells. Plant Cell Physiol. 42: 900-905.
		367634| Rn.24484	2.1E-16	[Rattus norvegicus] Member of the ion transport family, which contains calcium, sodium, and potassium channels, has weak similarity to a region of the alpha-1 subunit of voltage-gated calcium channels.
21	7503223CD1	g2739503	6.8E-80	[Homo sapiens] potassium channel
		g2739503	6.8E-80	[Homo sapiens] potassium channel. Su, K. et al. (1997) Isolation, characterization, and mapping of two human potassium Biochem. Biophys. Res. Commun. 241: 675-681.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
21 cont.		336122| KCNG1	6.0E-81	[Homo sapiens] [Channel (passive transporter); Transporter] [Plasma membrane; Unspecified membrane] Potassium voltage channel subfamily G 1, functions as a regulatory subunit, slows deactivation of the potassium channel upon association with Kv2.1 (KCNB1), may play a role in regulating membrane excitability. Su, K. et al. (1997) Isolation, characterization, and mapping of two human potassium channels. Biochem. Biophys. Res. Commun. 241: 675-81.
		428816| KCNG2	2.4E-24	[Homo sapiens] [Channel (passive transporter); Transporter] [Plasma membrane] Potassium voltage channel subfamily gamma 2, a member of the Kv6 family of ion channels, functions as a votage-gated potassium channel upon interaction with Kv2.1 alpha subunit, may contribute to cardiac action potentiation repolarization.
		430588| Kcnb1	1.2E-23	[Rattus norvegicus] [Channel (passive transporter); Transporter] [Plasma membrane; Unspecified membrane] Potassium voltage-gated channel (Shab-related subfamily, member 1), a delayed rectifier potassium channel which plays a role in vasoconstriction, hypertension, and probably regulates neuronal excitability; associated with type I diabetic cardiomyopathy.
		429486| Kcnb1	1.3E-23	[Mus musculus] [Channel (passive transporter); Transporter] [Plasma membrane; Unspecified membrane] Potassium voltage-gated channel (Shab-related subfamily, member 1), a delayed rectifier channel which plays a role in Schwann cell proliferation, glucose-stimulated insulin secretion, and probably regulates neuronal excitability.
		341660| KCNB1	1.3E-23	[Homo sapiens] [Channel (passive transporter); Transporter] [Plasma membrane; Unspecified membrane] Potassium voltage-gated channel (Shab-related subfamily, member 1), functions as a delayed rectifier potassium channel.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
22	7503566CD1	g1314290	2.0E-169	[Homo sapiens] vesicular monoamine transporter VMAT1. Erickson, J.D. et al.(1996) Distinct pharmacological properties and distribution in neurons and endocrine cells of two isoforms of the human vesicular monoamine transporter. Proc. Natl. Acad. Sci. U.S.A. 93: 5166-5171.
		338052| SLC18A1	1.7E-170	[Homo sapiens] Solute carrier family 18 member 1(vesicular monoamine transporter 1), transports and stores monoamine neurotransmitters in adrenal chromaffin cells, and is inhibited by the drugs fenfluramine, reserpine, and amphetamines.
		329252| Slc18a2	3.1E-165	[Rattus norvegicus] [Transporter] [Unspecified membrane; Other vesicles of the secretory/endocytic pathways] Solute carrier family18 A2 (vesicular monoamine transporter 2), pumps monoamines and neurotrasmitters into synaptic vesicles; human SLC18A2 genetic variants may be associated with neurobehavioral disorders.
		338054| SLC18A2	8.3E-165	[Homo sapiens] [Transporter] [Secretory vesicles; Cytoplasmic; Unspecified membrane; Plasma membrane; Otherv esicles of the secretory/endocytic pathways] Solute carrier family18 A2 (vesicular monoamine transporter 2), pumps monoamines and neurotransmitters into synaptic vesicles, activity is sensitive to inhibition by tetrabenazine; genetic variants may be associated with neurobehavioral disorders.
		331476| Slc18a1	1.2E-146	[Rattus norvegicus] [Transporter] [Secretory vesicles; Cytoplasmic; Unspecified membrane; Other vesicles of the secretory/endocytic pathways] Solute carrier family 18 member 1 (vesicular monoamine transporter 1), transports and stores monoamine neurotransmitters in adrenal chromaffin cells, and is inhibited by the drug reserpine.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
23	7505122CD1	g1085026	7.0E-23	[Homo sapiens] MAT8 protein. Morrison, B.W. et al. (1995) Mat-8, a novel phospholemman-like protein expressed in human breast tumors, induces a chloride conductance in Xenopus oocytes. J. Biol. Chem. 270 : 2176-2182.
		343156| FXYD3	6.1E-24	[Homo sapiens] [Regulatory subunit; Channel (passive transporter); Transporter] [Plasma membrane] FXYD domain containing ion transport regulator 3, channel or channel regulator that mediates chloride channel activity, expressed in breast cancers, potential marker of breast cancer progression; gene regulation varies in breast cancer cell lines. Schiemann, S. et al. (1998) Molecular analysis of two mammary carcinoma cell lines at the transcriptional level as a model system for progression of breast cancer. Clin. Exp. Metastasis 16:129-39.
		319786| Fxyd3	8.5E-18	[Mus musculus] [Channel (passive transporter);Transporter] [Plasma membrane] FXYD domain containing ion transport regulator 3, putative chloride channel or channel regulator; upregulated specifically in nue and ras-initiated mammary tumors indicating that the protein is involved in cellular differentiation and signaling.
		628631| LOC64190	1.2E-13	[Rattus norvegicus] [Regulatory subunit; Channel (passive transporter); Transporter] [Unspecified membrane] Channel-inducing factor, contains an extracytoplasmic FXYD domain, binds and regulates Na, K-ATPase in renal tissue, may play a role in maintenance of Na+ and K+ homeostasis; expression is induced by aldosterone.
24	7511620CD1	g13344999	4.0E-76	[Homo sapiens] solute carrier family 26 member 6
				Waldegger, S. et al., Cloning and characterization of slc26a6, a novel member of the solute carrier 26 gene family., Genomics 72, 43-50. (2001).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
24 cont.		662428| SLC26A6	1.6E-66	[Homo sapiens] Solute carrier family 26 member 6, a member of the anion transporter family that is present on the apical plasma membrane and may act in exchange of chloride and bicarbonate ions across membranes in an electroneutral manner.
				Lohi, H. et al., Mapping of five new putative anion transporter genes in human and characterization of SLC26A6, A candidate gene for pancreatic anion exchanger, Genomics 70, 102-12 (2000).
		430122| Slc26a4	1.8E-23	[Mus musculus][Transporter][Plasma membrane; Unspecified membrane] Solute carrier family 26 member 4 (pendrin), a putative bicarbonate transporter; mutation of the human SLC26A4 gene causes Pendred syndrome, enlarged vestibular aqueduct syndrome, and DFNB4 form of non syndromic hearing loss.
				Royaux, I. E. et al., Pendrin, encoded by the Pendred syndrome gene, resides in the apical region of renal intercalated cells and mediates bicarbonate secretion., Proc Natl Acad Sci U S A 98, 4221-6. (2001).
25	7506995CD1	g20384654	1.0E-129	[Homo sapiens] two-pore calcium channel protein 2
		g4586963	8.6E-27	[Rattus norvegicus] voltage-gated ca channel
				Ishibashi, K. et al.
				Molecular cloning of a novel form (Two-repeat) protein related to voltage-gated sodium and calcium channels
				Biochem. Biophys. Res. Commun. 270, 370-376 (2000)
		367634| Rn.24484	7.4E-28	[Rattus norvegicus] Member of the ion transport family, which contains calcium, sodium, and potassium channels, has weak similarity to a region of low voltage-activated T-type calcium channel alpha 1I subunit (rat Cacna1i).
				Ishibashi, K. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
25 cont.		605816| CACNA1G	7.0E-11	[Homo sapiens][Channel (passive transporter); Transporter][Plasma membrane; Unspecified membrane] Calcium channel (voltage-dependent, T-type) alpha 1G subunit, generates T-type calcium currents, mediates increases in intracellular calcium concentrations; may play a role in tumor development.
				Perez-Reyes, E. et al. Molecular characterization of a neuronal low-voltage-activated T-type calcium channel. Nature 391, 896-900 (1998).
				Toyota, M. et al. Inactivation of CACNA1G, a T-type calcium channel gene, by aberrant methylation of its 5' CpG island in human tumors. Cancer Res 59, 4535-41. (1999).
26	7506996CD1	g20384654	0.0	[Homo sapiens] two-pore calcium channel protein 2
		g7270679	5.2E-31	[Arabidopsis thaliana] putative calcium channel
		367634| Rn.24484	7.4E-32	[Rattus norvegicus] Member of the ion transport family, which contains calcium, sodium, and potassium channels, has weak similarity to a region of low voltage-activated T-type calcium channel alpha 1I subunit (rat Cacna1i) Ishibashi, K. et al. Biochem Biophys Res Commun 270, 370-6 (2000).

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1853191CD1	652	S77 S85 S127 S155 S210 S414 S509 S583 S607 S624 S638 T30 T90 T232 T276 T319 T587	N136 N151 N412 N502 N520	Cytosolic domain: M1-T32, T237-L240, R306-S325, T408-D532, S583-R652 Transmembrane domain: A33-A55, T214-I236, V241-Y263, C283-L305, A326-V348, I385-V407, F533-N555, F560-L582 Non-cytosolic domain: G56-D213, Y264-K282, A349-L384, Y556-A559	TMHMMER
2	7497369CD1	345	S89 S104 S144 S305 T124	N13	Cytosolic domain: M1-G128, D179-R184, A235-P240, E296-H345 Non-cytosolic domain:R147-T155, L208-W211, L264-G272 Transmembrane domain: L129-R146, F156-A178, P185-Y207, T212-F234, T241-P263, F273-P295	TMHMMER
3	1700438CD1	150	S125		Signal Peptide: M19-G36; M19-G38; M7-G36	HMMER
					Cytosolic domains: M1-M19, R77-L82 Transmembrane domains: V20-T42, I57-L76, V83-L105 Non-cytosolic domains: G43-S56, A106-G150	TMHMMER
					NADH-UBIQUINONE/PLASTOQUINONE OXIDOREDUCTASE CHAIN 6 PF00499: W12-Y32	BLIMPS_PFAM
4	535939CD1	537	S55 S175 S276	N90	Sodium/hydrogen exchanger family: I119-V520	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4 cont.			S380 S412 S477 T9 T62 T105 T474		Cytosolic domains: S107-G112, I162-W172, P230-W233, S329-T340, S412-T417, G513-V537 Transmembrane domains: V87-G106, G113-I135, T139-V161, S173-L192, L207-L229, G234-L256, G306-P328, F341-A374, I389-A411, V418-V440, V490-L512 Non-cytosolic domains: M1-R86, K136-P138, D193-R206, Q257-R305, G375-I388, C441-D489	TMHMMER
5	55118067CD1	1119	S64 S172 S187 S217 S222 S251 S396 S429 S545 S685 S965 T68 T123 T847		Transmembrane amino acid transporter protein: C30-S398	HMMER_PFAM
					Cytosolic domains: M1-N8, K59-K82, Q143-A148, T253-K322, K367-Q377 Transmembrane domains: W9-F31, I36-V58, M83-G105, V120-L142, S149-S171, F230-F252, A323-P340, T344-Y366, V378-V397 Non-cytosolic domains: K32-G35, D106-Q119, S172-I229, N341-E343, S398-S1119	TMHMMER
					NEUROFILAMENT; TRIPLET; DM04498|P12036|434-1019: V389-E877	BLAST_DOMO
6	7502087CD1	947	S26 S55 S105 S140 S145 S200 S283 S288 S458 S488 S670 S706 S740 S763 S777 S853 S861 S868 S886 S918 T13 T170 T202 T220 T301 T326 T363 T377	N218 N449 N510 N731	PAC motif: K93-L135	HMMER_PFAM
					PAS domain: I41-M60	HMMER_PFAM
					Cyclic nucleotide-binding domain: K609-A699	HMMER_PFAM
					Ion transport protein: P299-V511	HMMER_PFAM
					Cytosolic domains: D284-A399, T486-V491 Transmembrane domains: W261-S283, V400-I422, Y463-N485, F492-I514 Non-cytosolic domains: M1-D260, G423-K462, I515-H947	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6 cont.			T486 T522 T678		POTASSIUM CHANNEL IONIC CHANNEL PD118772: E702-S944	BLAST_PRODOM
					CHANNEL IONIC POTASSIUM K+ SUBUNIT HYPERPOLARIZATION-ACTIVATED PROTEIN PUTATIVE EAG LONG PD001039: S228-V339	BLAST_PRODOM
					CHANNEL PROTEIN IONIC POTASSIUM NONPHOTOTROPIC HYPOCOTYL PUTATIVE SUBUNIT REPEAT EAG PD009483: M1-L86	BLAST_PRODOM
					CHANNEL IONIC POTASSIUM SUBUNIT MERG1A RERG VENTRICULAR ERG K+ PUTATIVE PD009699: K174-L227	BLAST_PRODOM
					CAMP RECEPTOR PROTEIN CYCLIC NUCLEOTIDE-BINDING DOMAIN DM01165|I38465|562-948: H413-F727 DM01165|I48912|391-786: H413-R715 DM01165|Q02280|384-776: H413-R715	BLAST_DOMO
					POTASSIUM; CHANNEL; KST1; AKT1; DM02383|I38465|353-560: T201-A412	BLAST_DOMO
7	7500819CD1	80	S19 S58 T36 Y32		signal_cleavage: M1-A18	SPSCAN
					Signal Peptides: M1-A18, M1-A20, M1-A21, M1-K25, M1-K23	HMMER
					ATP1G1/PLM/MAT8 family: A21-C75	HMMER_PFAM
					Cytosolic domains: M1-E2, R59-Q80 Transmembrane domains: L3-A21, T36-S58 Non-cytosolic domain: E22-Q35	TMHMMER
					ATP1G1 / PLM / MAT8 family proteins BL01310: D28-C63	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7 cont.					PHOSPHOLEMMAN PRECURSOR TRANSMEMBRANE PHOSPHORYLATION SIGNAL IONIC CHANNEL ION TRANSPORT PD058885: L5-F65	BLAST_PRODOM
					TRANSMEMBRANE TRANSPORT PRECURSOR SIGNAL CHANNEL ION SODIUM/POTASSIUM-TRANSPORTING ATPASE GAMMA CHAIN PD005989: D28-P69	BLAST_PRODOM
					CAMP; MAJOR; SODIUM; TRANSMEMBRANE; DM02637|A40533|19-92: A21-F65	BLAST_DOMO
					ATP1G1 / PLM / MAT8 family signature: D28-G41	MOTIFS
8	7503413CD1	531	S291 S313 T527	N525	Signal Peptide: M1-A33, M1-C38	HMMER
					Sodium: sulfate symporter transmembrane: S6-W341, P356-L517	HMMER_PFAM
					Cytosolic domains: M1-K12, P68-T86, N141-L210, Y276-K308, T423-N428, G481-T491 Transmembrane domains: S13-L30, M45-F67, N87-H106, P121-S140, C211-L233, W253-V275, L309-F331, L403-F422, V429-L448, T458-Y480, G492-I514 Non-cytosolic domains: M31-L44, K107-K120, L234-S252, S332-P402, N449-C457, F515-T531	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8 cont.					Sodium: sulfate symporter family proteins BL01271: T132-I151, M208-V232, P370-G391, S443-I497	BLIMPS_BLOCKS (P< 1.3E-10)
					PROTEIN TRANSMEMBRANE TRANSPORT MEMBRANE INNER TRANSPORTER SODIUM SYMPORT OF COTRANSPORTER PD000549: M208-W510, V15-V173	BLAST_PRODOM
					RENAL; BOUND; PRO-SER-ALA; NA; DM02914 P46556|1-520: R37-W341, L320-F515 A47714|28-576: P358-F515, T144-W341, I29-K206 S43561|28-507: R37-M159, L359-T509, V152-W341, A4-V41 P32739|25-517: G336-F515, M125-W341, R37-M159, P21-T86	BLAST_DOMO
					Na_Sulfate symporter: A461-V477	MOTIFS
9	7500007CD1	510	S10 T138 T433 T463 T495 T504	N77 N81 N86 N177	Sodium: neurotransmitter symporter family: E68-P478, R32-G67	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9 cont.					Cytosolic domains: K133-K140, K213-A224, D297-E325, A381-P398, L464-C510 Transmembrane domains: G110-W132, V141-L163, A190-N212, L225-M247, V274-L296, V326-M348, A358-G380, L399-I421, L441-T463 Non-cytosolic domains: M1-L109, P164-D189, S248-A273, Y349-Y357, K422-A440	TMHMMER
					Sodium: neurotransmitter symporter family proteins BL00610: W114-G165, R180-T232, A273-P315, S372-L426, Y432-P454	BLIMPS_BLOCKS
					Sodium: neurotransmitter symporter family signatures:N36-F129	PROFILESCAN
					SODIUM/NEUROTRANSMITTER PR00176: V141-I158, I223-I243, L277-L296, G361-A381, E40-L61, K401-I421	BLIMPS_PRINTS
					TRANSPORTER NEUROTRANSMITTER TRANSPORT TRANSMEMBRANE SYMPORT GLYCOPROTEIN SODIUM CHLORIDE-DEPENDENT SODIUM DEPENDENT GABA PD000448: G134-P478, N54-Y150, R32-G67	BLAST_PRODOM
					TRANSMEMBRANE TRANSPORT PROTEIN TRANSPORTER AMINO ACID PERMEASE AMINO ACID GLYCOPROTEIN MEMBRANE PD000214: A166-P454, A111-S261	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9 cont.					SODIUM CHLORIDE-DEPENDENT GABA TRANSPORTER NEUROTRANSMITTER TRANSPORT TRANSMEMBRANE GLYCOPROTEIN SYMPORT MULTIGENE PD037836: T463-C510	BLAST_PRODOM
					SODIUM: NEUROTRANSMITTER SYMPORTER FAMILY DM00572 P31646|22-576:E68-R485, E22-G67 P31649|22-576:E68-R485, E22-G67 P27799|26-581:E68-Q484, E22-G67 P48066|40-596:T8-R485, K24-G67	BLAST_DOMO
10	7500025CD1	894	S2 S26 S55 S105 S140 S145 S200 S283 S288 S435 S617 S653 S687 S710 S724 S800 S808 S815 S833 S865 T13 T170 T202 T220 T301 T326 T363 T377 T433 T469 T625	N218 N457 N678	PAC motif: K93-L135	HMMER_PFAM
					Cyclic nucleotide-binding domain: K556-A646	HMMER_PFAM
					Ion transport protein: P299-V458	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10 cont.					Cytosolic domains: M1-A257, R320-A338, C419-V438 Transmembrane domains: V258-F280, L300-F319, V339-F361, Y396-I418, F439-I461 Non-cytosolic domains: L281-P299, R362-E395, I462-H894	TMHMMER
					POTASSIUM CHANNEL IONIC CHANNEL PD118772: E649-S891	BLAST_PRODOM
					CHANNEL IONIC POTASSIUM K+ SUBUNIT HYPERPOLARIZATION ACTIVATED PROTEIN PUTATIVE EAG LONG PD001039: S228-V339	BLAST_PRODOM
					CHANNEL PROTEIN IONIC POTASSIUM NON-PHOTOTROPIC HYPOCOTYL PUTATIVE SUBUNIT REPEAT EAG PD009483: R4-L86	BLAST_PRODOM
					CHANNEL IONIC POTASSIUM SUBUNIT MERG1A RERG VENTRICULAR ERG K+ PUTATIVE PD009699: K174-L227	BLAST_PRODOM
					CAMP RECEPTOR PROTEIN CYCLIC NUCLEOTIDE-BINDING DOMAIN DM01165 I38465|562-948: S420-F674, H413-I418 Q02280|384-776: A412-R662, H413-I418 I48912|391-786: S420-R662, H413-I418	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10 cont.					POTASSIUM; CHANNEL; KST1; AKT1; DM02383|I38465|353-560: T201-A412	BLAST_DOMO
11	7502736CD1	788	S19 S24 S47 S60 S64 S79 S85 S124 S129 S299 S329 S511 S547 S581 S604 S618 S694 S702 S709 S727 S759 T33 T142 T167 T204 T218 T327 T363 T519	N290 N351 N572	Cyclic nucleotide-binding domain: K450-A540	HMMER_PFAM
					Ion transport protein: P140-V352	HMMER_PFAM
					Cytosolic domains: M1-D101, N159-A240, T327-V332 Transmembrane domains: W102-S124, S139-I158, V241-I263, Y304-N326, F333-I355 Non-cytosolic domains: D125-C138, G264-K303, I356-H788	TMHMMER
					POTASSIUM CHANNEL IONIC CHANNEL PD118772: E543-S785	BLAST_PRODOM
					CHANNEL IONIC POTASSIUM K+ SUBUNIT HYPERPOLARIZATIONACTIVATED PROTEIN PUTATIVE EAG LONG PD001039: A94-V180	BLAST_PRODOM
					CAMP RECEPTOR PROTEIN CYCLIC NUCLEOTIDE-BINDING DOMAIN DM01165|I38465|562-948: H254-F568 DM01165|I48912|391-786: H254-R556 DM01165|Q02280|384-776: H254-R556	BLAST_DOMO
					POTASSIUM; CHANNEL; KST1; AKT1; DM02383|I38465|353-560: A94-A253	BLAST_DOMO
12	7503570CD1	501	S64 S152 S185 S195 S200 S289 S293 S309 S366	N264 N280 N323	Cytosolic domain: M1-R81 Transmembrane domain: W82-V104 Non-cytosolic domain: R105-A501	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12 cont.			S398 S478 T5 T220		HEAVY CHAIN CD98 PROTEIN 4F2 CELL SURFACE ANTIGEN 4F2HC GLYCOPROTEIN TRANSMEMBRANE PD013690: Q221-P387, L373-A501 PD011378: M1-R125	BLAST_PRODOM
					CELL SURFACE ANTIGEN 4F2 HEAVY CHAIN DM08844|P08195|113-428: P113-Q381, S366-R401 DM08844|P10852|107-423: P113-Q381, Q381-R401	BLAST_DOMO
					1LAT; CYCLOMALTODEXTRINASE; DM04732|P08195|430-528: S402-A501 DM04732|S57719|424-527: R401-Y499	BLAST_DOMO
13	7504008CD1	721	S156 S338 S536 S565 S578 S596 S641 S650 S676 S696 T138 T174 T251 T577 T619 T689 Y666	N167 N172 N576	STAS domain: Y493-A700	HMMER_PFAM
					Sulfate transporter family: L198-T470	HMMER_PFAM
					Cytosolic domains: T138-T186, R331-E341, F399-K404, R469-L721 Transmembrane domains: A115-G137, L187-Y209, G308-L330, L342-M364, V379-L398, A405-L422, L437-V468 Non-cytosolic domains: M1-L114, L210-V307, S365-Q378, S423-L436	TMHMMER
					Sulfate transporters proteins BL01130: S96-V149, T186-L237	BLIMPS_BLOCKS
					SULFATE TRANSPORTER TRANSPORT PROTEIN TRANSMEMBRANE GLYCOPROTEIN AFFINITY SULPHATE HIGH PERMEASE PD001121: L71-V196	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13 cont.					PROTEIN TRANSPORT SULFATE TRANSPORTER TRANSMEMBRANE PERMEASE INTERGENIC REGION AFFINITY GLYCOPROTEIN PD001255: H278-R469, L198-L237	BLAST_PRODOM
					SULFATE TRANSPORTER PROTEIN TRANSPORT TRANSMEMBRANE AFFINITY GLYCOPROTEIN SULPHATE HIGH DISEASE PD001755: P473-L523, Q627-R709	BLAST_PRODOM
					SULFATE TRANSPORTERS DM01229|P45380|10-468: L264-K430, C62-A175, E159-L264, L422-L478 DM01229|P50443|49-505: L264-W429, Q61-L187, S160-L264 DM01229|P40879|5-462: Y29-G317, A267-K430 DM01229|Q02920|1-447: R66-W257, P260-K430	BLAST_DOMO
14	7503559CD1	1226	S98 S103 S130 S144 S170 S227 S252 S523 S802 S817 S899 S901 S1055 T74 T269 T353 T358 T387 T502 T549 T576 T912 T1061 T1159 T1182 T1206 Y349 Y407	N23 N150 N300 N312 N318 N704 N1045 N1053 N1059 N1073 N1217	E1-E2 ATPase: N278-V365	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14 cont.					Cytosolic domains: M1-K28, R254-Y411, A469-A930, N986-A1005, K1093-N1104, E1161-T1226 Transmembrane domains: L29-P51, Y231-I253, L412-L434, I449-Y468, A931-V953, F968-L985, L1006-V1028, T1075-S1092, Y1105-A1127, W1142-V1160 Non-cytosolic domains: E52-E230, N435-D448, T954-Q967, K1029-T1074, S1128-Q1141	TMHMMER
					E1-E2 ATPases phosphorylation site proteins BL00154: V454-G490, L492-L510, K652-C662, N724-M764, V878-S901, A905-V938	BLIMPS_BLOCKS
					E1-E2 ATPases phosphorylation site: I478-E526	PROFILESCAN
					P-type cation-transporting atpase superfamily signature PR00119: N318-T332, C496-L510, A740-D750, C881-L900	BLIMPS_PRINTS
					ATPASE PROBABLE CALCIUM-TRANSPORTING PROTEIN HYDROLASE CALCIUM TRANSPORT TRANSMEMBRANE PHOSPHORYLATION MAGNESIUM PD090368: D1064-L1114, Q995-Y1034; PD150086: G178-L236; PD023991: E904-I1016, E1062-D1165	BLAST_PRODOM
					ATPASE HYDROLASE TRANSMEMBRANE PHOSPHORYLATION ATP-BINDING TRANSPORT PUMP CALCIUM MAGNESIUM MEMBRANE PD000132: E282-D506, Q672-I765, D715-M764	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14 cont.					E1-E2 ATPASES PHOSPHORYLATION SITE DM00115|P22189|49-801: V623-P767, H800-S984, S202-K331, P401-E505, S556-A575 DM00115|P13586|78-773: S202-N527, E840-D972, Q607-H790, G557-H583 DM00115|P54678|80-795: I622-I765, A235-D506, M823-Y957, V1132-D1165 DM00115|P47317|26-695: F628-K780, I238-E566, D844-L963	BLAST_DOMO
					E1-E2 ATPases phosphorylation site: D498-T504	MOTIFS
15	6243872CD1	638	S11 S41 S75 S129 S263 S366 S371	N215 N221 N418 N465 N580	Organic Anion Transporter Polypeptide (OATP): G156-I394, N465-L638	HMMER_PFAM
					Cytosolic domains: M1-C107, D176-Y187, N258-V341, R397-P498, Y610-L638 Transmembrane domains: F108-I130, I153-I175, L188-V210, W235-P257, A342-I364, A374-V396, L499-A521, L587-I609 Non-cytosolic domains: G131-D152, K211-T234, V365-K373, M522-F586	TMHMMER
					TRANSPORTER PROTEIN TRANSMEMBRANE TRANSPORT SIMILAR MATRIN F/G ORGANIC ANION SODIUMINDEPENDENT PD005488: C423-T618	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15 cont.					ANION; ORGANIC; TRANSPORTER; INDEPENDENT; DM03420|Q00910|1-642: G161-T618, N103-V159 DM03420|P46721|1-638: G156-I591 DM03420|A49580|1-638: G156-I591 DM03420|P46720|1-638: G156-I591	BLAST_DOMO
16	90011608CD1	507	S4 S46 S51 S55 S143 S365 T34	N62 N76 N141 N205 N214 N256	Transmembrane amino acid transporter protein: A102-G482	HMMER_PFAM
					Cytosolic domains: K131-R171, R243-L248, L309-P320, S395-D406, S457-D468 Transmembrane domains: I108-V130, R172-A194, S225-I242, S249-V271, W286-P308, L321-L343, I372-V394, L407-I424, V434-Y456, A469-I491 Non-cytosolic domains: M1-G107, D195-D224, Q272-P285, Q344-S371, P425-L433, Q492-I507	TMHMMER
					ACID AMINO PROTEIN TRANSPORTER PERMEASE TRANSMEMBRANE INTERGENIC REGION PUTATIVE PROLINE PD001875: W80-L380	BLAST_PRODOM
					PROTEIN F59B2.2 Y43F4B.7 H32K16.1 SIMILAR C ELEGANS CHROMOSOME III PD150950: V174-L333	BLAST_PRODOM
					PROTEIN F59B2.2 T27A1.5 Y43F4B.7 H32K16.1 SIMILAR C ELEGANS CHROMOSOME III PD150568: V328-Y456	BLAST_PRODOM
					PROTEIN Y43F4B.7 H32K16.1 SIMILAR C ELEGANS F59B2.2 PD018165: N76-R164	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	90024583CD1	568	S291 S313 S372 S381 T383 T564	N562	Signal Peptide: M1-A33, M1-C38	HMMER
					Sodium: sulfate symporter transmembrane: S6-L554	HMMER_PFAM
					Cytosolic domains: M1-K12, P68-N87, Q157-L210, Y276-K308, P371-G409, N465-G484, N545-T568 Transmembrane domains: S13-L30, M45-F67, M88-L105, P134-L156, C211-L233, W253-V275, L309-F331, Y353-V370, I410-W432, A442-S464, L485-A507, V522-V544 Non-cytosolic domains: M31-L44, H106-A133, L234-S252, S332-K352, M433-H441, T508-K521	TMHMMER
					Sodium: sulfate symporter family proteins BL01271: T132-I151, M208-V232, P407-G428, S480-I534	BLIMPS_BLOCKS
					PROTEIN TRANSMEMBRANE TRANSPORT MEMBRANE INNER TRANSPORTER SODIUM SYMPORT OF COTRANSPORTER PD000549: F323-W547, V15-V173, M208-K374, I16-L170	BLAST_PRODOM
					RENAL; BOUND; PRO-SER-ALA; NA; DM02914|P46556|1-520: R37-F552 DM02914|A47714|28-576: T144-F552, I29-K206 DM02914|S43561|28-507: E290-T546, R37-M159, V152-K388, A4-V41 DM02914|P32739|25-517: D174-F552, R37-M159, P21-T86	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17 cont.					Sodium: sulfate symporter family signature: A498-V514	MOTIFS
18	90113658CD1	595	S84 S96 S105 S126 S275 S287 S319 S440 S540 T52 T194 T296 T563 T582	N60 N79 N82 N103 N188 N538	Cytosolic domains: M1-S144, M191-R196, L247-Q252, R401-G404, E470-V475, P528-M595 Transmembrane domains: L145-G164, V168-T190, F197-R214, F224-A246, V253-F272, Y378-G400, L405-G427, I447-A469, I476-I498, L508-L527 Non-cytosolic domains: R165-P167, I215-R223, P273-D377, K428-S446, E499-F507	TMHMMER
					Sugar transport proteins signature 1: T156-S172	MOTIFS
19	3942766CD1	602	S13 S78 S101 S220 S305 S394 S398 T318 T410 T417 T575 T598	N312 N586 N596	Signal Peptide: M1-A33	HMMER
					Sodium: sulfate symporter transmembrane: P213-L573, A7-L162	HMMER_PFAM
					Cytosolic domains: M1-V11, P77-K80, K160-L235, Y301-N334, P393-A467, P528-T547 Transmembrane domains: W12-L34, T54-L76, V81-I103, F137-L159, I236-I258, W278-L300, L335-F357, L370-F392, V468-I490, H505-T527, G548-I570 Non-cytosolic domains: P35-C53, E104-S136, L259-S277, T358-S369, F491-V504, F571-L602	TMHMMER
					Sodium: sulfate symporter family proteins BL01271: T135-I154, F234-I258, P426-G447, A499-L553	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19 cont.					PROTEIN TRANSMEMBRANE TRANSPORT MEMBRANE INNER TRANSPORTER SODIUM SYMPORT OF CO-TRANSPORTER PD000549: I328-W566, L17-L162, L235-W402	BLAST_PRODOM
					RENAL; BOUND; PRO-SER-ALA; NA; DM02914 |A47714|28-576: A562-F571 |S43561|28-507: G233-A567, E38-L162, L549-N596 |P46556|1-520: K204-F571, E38-E177 |P32739|25-517: K232-F571, E38-A182, L21-A63	BLAST_DOMO
20	7501987CD1	372	S24 S29 S162 S180		Cytosolic domains: M1-A77, D244-D254 Transmembrane domains: I78-Y100, Y221-L243, F255-L274 Non-cytosolic domains: S101-Y220, K275-V372	TMHMMER
21	7503223CD1	165	S9 S146 T79 T80 T96 T111		K+ channel tetramerisation domain: R63-S159	HMMER_PFAM
					POTASSIUM CHANNEL IONIC CHANNEL PD060713: M1-R62	BLAST_PRODOM
					CHANNEL; POTASSIUM; CDRK; SHAW; DM00490 |JH0595|26-142: R63-D151 |P15387|18-134: R63-D151 |P17970|268-384: R64-D149 |P17972|1-102: D60-S159	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
23 cont.					Cytosolic domain: S57-F67 Transmembrane domain: W34-T56 Non-cytosolic domain: M1-D33	TMHMMER
					Copper/Zinc superoxide dismutase signatures: L12-R60	PROFILESCAN
					Nickel-dependent hydrogenases large subunit signatures: G7-K59	PROFILESCAN
					ATP1G1/PLM/MAT8 family BL01310: S28-L63	BLIMPS_BLOCKS
					CAMP; MAJOR; SODIUM; TRANSMEMBRANE; DM02637 |A55571|20-87: A20-G52 |I48648|20-88: A20-G52	BLAST_DOMO
					ATP1G1 / PLM / MAT8 family signature: S28-G41	MOTIFS
24	7511620CD1	152	T138		Cytosolic domain: T138-Q152 Transmembrane domain: A115-G137 Non-cytosolic domain: M1-L114	TMHMMER
					Sulfate transporter IPB001902: L105-Q152, L75-Y84	BLIMPS_BLOCKS
					SULFATE TRANSPORTER, TRANSPORT PROTEIN, TRANSMEMBRANE GLYCOPROTEIN, HIGH AFFINITY SULPHATE, PERMEASE PD001121: L71-G145	BLAST_PRODOM
					SULFATE TRANSPORTERS DM01229 |P50443|49-505: Q61-L146 |P40879|5-462: Y29-G145 |P45380|10-468: C62-L146 |P38359|53-515: H54-G145	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
22	7503566CD1	497	S259 S305 S489 T4 T136 T161 T476	N58 N87 N104	Cytosolic domains: M1-L22, T192-T203, S253-P272, N333-W338, D391-H394, L450-E497 Transmembrane domains: V23-V45, P169-Y191, A204-V226, A230-P252, Y273-L295, L310-A332, L339-I361, L371-V390, T395-P417, I427-Y449 Non-cytosolic domains: P46-I168, G227-S229, P296-Q309, F362-G370, S418-A426	TMHMMER
					CHROMAFFIN GRANULE AMINE TRANSPORTER VESICULAR VAT1 TRANSPORT TRANSMEMBRANE GLYCOPROTEIN NEUROTRANSMITTER PD034226: M1-G128	BLAST_PRODOM
					CHROMAFFIN GRANULE AMINE TRANSPORTER VESICULAR 1 VAT1 TRANSPORT TRANSMEMBRANE GLYCOPROTEIN NEUROTRANSMITTER PD034225: P217-F266 PD164279: A461-E497	BLAST_PRODOM
					SIMILAR TO SYNAPTIC VESICLE AMINE TRANSPORTER PD170805: M186-D271	BLAST_PRODOM
					VESICULAR LUMEN DOMAIN DM03018 |P54219|172-524: F178-E497 |S43685|140-492: F178-G486 |I38658|157-516: F178-S452 |P34711|135-500: Y166-R451	BLAST_DOMO
23	7505122CD1	67	S57 S62 T56		signal_cleavage: M1-A20	SPSCAN
					Signal Peptide: M1-A20	HMMER
					ATP1G1/PLM/MAT8 family: A20-F67	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25	7506995CD1	467	S24 S29 S150 S254 S276 S381 T143 T454	N326 N333	Cytosolic domains: M1-A77, N293-G298, D375-K386 Transmembrane domains: I78-Y100, F270-Q292, G299-V321, A355-L374, I387-L409 Non-cytosolic domains: S101-R269, A322-F354, E410-R467	TMHMMER
26	7506996CD1	490	S24 S29 S162 S180 S404 T477	N349 N356	Cytosolic domains: M1-A77, D244-D254, Q315-F320, D398-K409 Transmembrane domains: I78-Y100, Y221-L243, F255-F277, F292-V314, G321-I343, A378-L397, I410-L432 Non-cytosolic domains: S101-Y220, A278-V291, V344-F377, E433-R490	TMHMMER

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
27/1853191CB1/ 2343	1-296, 1-497, 1-528, 192-603, 329-565, 471-575, 471-1030, 493-822, 558-1165, 650-889, 650-907, 695-1152, 696-976, 754-1428, 881-1454, 885-1520, 891-1172, 891-1374, 892-1154, 917-1530, 918-1557, 931-1141, 942-1294, 960-1492, 965-1415, 997-1223, 1020-1634, 1074-1686, 1074-1708, 1077-1289, 1077-1543, 1094-1644, 1126-1668, 1129-1294, 1129-1451, 1169-1766, 1193-1389, 1244-1685, 1275-1895, 1279-1809, 1290-1599, 1295-1451, 1295-1535, 1295-1538, 1295-1824, 1305-1823, 1312-1847, 1314-1554, 1451-1705, 1452-1673, 1492-1774, 1533-1753, 1533-2102, 1536-1673, 1569-1833, 1569-2010, 1602-1775, 1602-2145, 1616-1845, 1616-2018, 1616-2146, 1642-1879, 1649-1860, 1650-1871, 1650-1898, 1650-2141, 1663-2145, 1669-2108, 1670-2102, 1674-2112, 1679-2109, 1701-1982, 1718-2146, 1728-2295, 1736-2144, 1743-2013, 1760-2103, 1762-2132, 1792-2007, 1792-2104, 1796-2143, 1797-2089, 1802-2058, 1820-2062, 1826-2145, 1876-2145, 1898-2146, 1902-2286, 1905-2114, 1997-2145, 2037-2343
28/7497369CB1/ 3145	1-772, 3-2492, 6-767, 141-767, 195-772, 308-772, 352-772, 439-982, 804-1160, 857-1118, 857-1265, 857-1270, 857-1285, 857-1413, 857-1548, 952-1118, 971-1341, 988-1597, 1051-1660, 1087-1245, 1087-1251, 1343-1629, 1361-2079, 1478-1665, 1478-1930, 1512-1991, 1639-1904, 1642-1944, 1708-2211, 1786-2413, 1803-2023, 1803-2208, 1821-2389, 1886-2329, 1971-2270, 1971-2373, 1984-2482, 1994-2481, 2018-2476, 2110-2492, 2215-2460, 2215-2478, 2215-2548, 2238-2473, 2291-2492, 2317-2498, 2363-3145, 2368-2526
29/1700438CB1/ 763	1-521, 2-763, 337-763, 397-627, 505-763, 546-761, 548-763
30/535939CB1/ 2720	1-702, 174-780, 183-666, 276-961, 366-632, 395-866, 437-683, 437-942, 454-864, 578-1039, 613-1199, 616-937, 617-852, 617-855, 632-1052, 659-1152, 740-1312, 805-997, 822-1098, 835-1471, 835-1478, 975-1563, 977-1565, 1041-1650, 1043-1529, 1066-1597, 1070-1611, 1072-1217, 1080-1311, 1080-1634, 1153-1577, 1182-1812, 1200-1505, 1214-1872, 1249-1523, 1302-1587, 1312-1815, 1460-1907, 1460-2081, 1468-1691, 1468-1701, 1475-1727, 1478-2055, 1491-2070, 1510-2132, 1514-2076, 1551-1830, 1570-1846, 1603-2219, 1620-1839, 1620-1891, 1620-2090, 1620-2131, 1620-2198, 1623-2090, 1627-2250, 1635-2022, 1645-1846, 1690-2109, 1691-2232, 1701-2306, 1705-2090, 1735-2028, 1810-2028, 1810-2038, 1810-2205, 1838-2047, 1838-2313, 1859-2091, 1912-2256, 1995-2518, 2053-2442, 2053-2648, 2070-2645, 2074-2232, 2074-2631, 2074-2720, 2137-2403, 2144-2702, 2167-2319, 2175-2666, 2191-2720, 2223-2514, 2223-2682, 2223-2702, 2233-2693, 2237-2718, 2240-2720, 2290-2716, 2299-2671, 2321-2677, 2323-2598, 2354-2555, 2354-2681, 2376-2714, 2379-2716, 2381-2716, 2472-2704, 2472-2718, 2488-2712, 2513-2718, 2550-2716, 2628-2720

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
31/55118067CB1/ 4464	1-703, 1-727, 20-559, 20-783, 39-617, 56-792, 137-724, 257-942, 373-747, 373-857, 375-843, 379-909, 381-567, 381-776, 381-867, 381-964, 381-971, 381-1047, 382-805, 384-728, 401-846, 415-723, 428-1040, 436-760, 436-1011, 450-574, 451-1136, 466-953, 484-870, 574-1183, 629-1210, 636-1265, 649-1211, 696-1124, 696-1129, 696-1149, 696-1160, 696-1180, 696-1198, 696-1261, 696-1290, 696-1299, 696-1304, 696-1312, 696-1488, 696-1519, 699-1542, 711-1265, 725-1325, 758-1432, 763-1326, 780-1055, 780-1199, 780-1234, 780-1295, 780-1397, 780-1436, 780-1439, 780-1450, 780-1453, 780-1465, 780-1470, 780-1480, 780-1520, 792-1589, 801-1486, 830-1645, 851-1564, 861-1102, 861-1352, 861-1385, 861-1386, 861-1405, 861-1429, 861-1544, 862-1477, 866-1592, 869-1723, 893-1486, 912-1138, 921-1744, 932-1136, 939-1667, 943-1564, 945-1564, 946-1132, 947-1486, 947-1561, 1011-1592, 1012-1564, 1017-1966, 1038-1514, 1119-1777, 1124-1718, 1143-1692, 1218-1665, 1472-2153, 1510-2187, 1529-2226, 1563-2240, 1581-2047, 1598-2109, 1605-1968, 1615-2263, 1624-2359, 1626-2099, 1646-2305, 1655-2242, 1673-2298, 1700-2212, 1725-2236, 1755-2195, 1756-2005, 1765-2471, 1777-2362, 1792-2467,
	1892-2416, 1980-2263, 2010-2284, 2017-2652, 2045-2589, 2121-2286, 2230-2620, 2234-3009, 2244-2513, 2248-2487, 2248-2601, 2431-2455, 2445-3149, 2666-3207, 2672-3315, 2677-3362, 2682-3216, 2692-3279, 2740-3429, 2787-3449, 2843-3545, 2854-3331, 2873-3567, 2883-3364, 2944-3162, 2950-3154, 2961-3289, 2980-3593, 2982-3268, 2990-3615, 3006-3251, 3011-3340, 3034-3707, 3041-3303, 3090-3358, 3090-3423, 3090-3854, 3099-3335, 3099-3364, 3114-3332, 3120-3574, 3139-3358, 3152-3415, 3163-3424, 3286-3924, 3304-3515, 3308-3513, 3378-4239, 3381-3949, 3381-3963, 3400-3872, 3405-3661, 3417-4018, 3489-3687, 3489-3711, 3492-4196, 3524-4243, 3527-4130, 3538-4206, 3538-4208, 3546-4071, 3550-4112, 3554-4210, 3556-4195, 3562-4182, 3570-3855, 3590-4266, 3592-4003, 3604-3813, 3609-3874, 3612-4091, 3633-4215, 3636-4229, 3638-3916, 3653-4238, 3664-4270, 3669-4278, 3673-4268, 3683-4223, 3690-4229, 3692-3905, 3692-4464, 3694-4239, 3697-4166, 3703-4266, 3720-4192, 3730-3880, 3730-3984, 3731-4079, 3741-3997, 3744-4017, 3745-4022, 3752-4006, 3777-4000, 3798-4062, 3798-4246, 3847-4263, 3873-4239, 3873-4242, 3926-4180, 4027-4227

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
32/7502087CB1/ 3135	1-676, 1-761, 8-676, 11-676, 12-315, 12-661, 12-676, 14-676, 15-676, 16-676, 17-676, 18-676, 24-676, 25-699, 25-706, 25-716, 25-743, 44-676, 63-676, 66-676, 622-1510, 664-1483, 680-1476, 924-1736, 1456-2345, 1657-2241, 1663-2628, 1683-2425, 1706-2496, 1716-2657, 1717-2657, 1720-2657, 1722-2655, 1729-2433, 1738-2571, 1742-2434, 1747-2657, 1753-2633, 1760-2365, 1770-2657, 1774-2657, 1775-2477, 1776-2310, 1811-2654, 1813-2562, 1815-2655, 1826-2657, 1830-2657, 1866-2657, 1867-2657, 1872-2657, 1876-2657, 1882-2657, 1884-2653, 1888-2657, 1895-2657, 1911-2657, 1912-2657, 1918-2657, 1923-2657, 1926-2657, 1945-2657, 1948-2657, 1953-2657, 1957-2657, 1962-2657, 1963-2657, 1969-2657, 1975-2657, 1979-2657, 1988-2657, 1992-2329, 1993-2657, 2005-2653, 2015-2408, 2016-2657, 2021-2657, 2022-2657, 2025-2657, 2033-2657, 2037-2657, 2050-2657, 2089-2657, 2097-2657, 2104-2657, 2115-2657, 2125-2657, 2127-2657, 2140-2657, 2144-2657, 2151-2657, 2159-2657, 2165-2657, 2174-2572, 2179-2657, 2180-2577, 2200-2654, 2202-2657, 2206-2657, 2217-2630, 2243-2657, 2250-2792, 2259-2657, 2298-2794, 2330-2526, 2348-2846, 2348-2893, 2352-2929, 2364-2929, 2374-2606, 2612-3135
33/7500819CB1/ 843	1-128, 1-133, 1-157, 1-170, 1-200, 1-201, 1-202, 1-208, 1-213, 1-215, 1-217, 1-218, 1-223, 1-234, 1-251, 1-255, 1-798, 4-209, 43-193, 43-208, 47-181, 48-208, 80-537, 111-208, 254-818, 261-828, 265-686, 268-815, 274-790, 279-572, 279-797, 280-598, 283-537, 298-523, 302-815, 313-536, 313-795, 331-799, 337-593, 347-796, 351-817, 353-834, 356-796, 356-813, 364-797, 366-609, 367-814, 369-802, 381-797, 386-802, 394-801, 395-795, 395-801, 401-797, 401-800, 404-631, 404-797, 405-832, 412-798, 413-698, 417-669, 417-797, 420-800, 423-800, 424-795, 425-798, 427-572, 428-800, 428-802, 433-795, 438-796, 450-748, 452-795, 464-804, 471-808, 484-797, 484-798, 485-832, 486-710, 486-795, 486-799, 489-798, 494-797, 504-795, 510-797, 524-797, 526-843, 533-756, 554-798, 565-766, 567-697, 622-819, 650-751, 651-842, 652-797, 652-843, 656-795, 677-787

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
34/7503413CB1/ 3159	1-701, 1-812, 1-3090, 6-869, 26-744, 444-842, 444-939, 444-994, 444-1006, 444-1010, 444-1047, 707-1050, 1055-1549, 1503-1693, 1537-2137, 1544-2109, 1570-1797, 1668-1916, 1749-2223, 1752-2422, 1772-2390, 1793-2434, 1818-2157, 1818-2212, 1819-2304, 1879-3087, 1883-2130, 1883-2148, 1883-2304, 1883-2349, 1883-2351, 1883-2362, 1883-2368, 1883-2390, 1883-2411, 1883-2420, 1883-2456, 1883-2463, 1883-2551, 1884-2179, 1885-2454, 1888-2545, 1888-2690, 1889-2541, 1945-2059, 1952-2629, 1976-2519, 1982-2395, 2079-2627, 2091-2531, 2091-2568, 2091-2584, 2095-2331, 2095-2667, 2112-2568, 2120-2669, 2131-2716, 2178-2647, 2189-2723, 2233-2669, 2238-2877, 2246-2750, 2265-2578, 2272-2760, 2287-3096, 2298-2806, 2351-2976, 2351-3052, 2362-2809, 2372-3043, 2407-3112, 2412-3090, 2422-3085, 2459-3077, 2478-2992, 2486-3123, 2512-3119, 2517-3118, 2518-3101, 2534-3034, 2542-3035, 2543-2802, 2543-2975, 2543-3001, 2543-3067, 2543-3091, 2543-3100, 2543-3101, 2543-3102, 2593-3079, 2593-3116, 2607-3107, 2615-3092, 2615-3093, 2626-3078, 2629-3092, 2629-3098, 2630-3085, 2637-3088, 2639-3090, 2639-3133, 2660-2921, 2660-3047, 2660-3087, 2680-3086, 2680-3117, 2681-3087,
	2684-2939, 2684-3087, 2690-3159, 2691-3104, 2714-2944, 2805-3102, 2901-3090
35/7500007CB1/ 1883	1-245, 12-222, 12-575, 12-578, 12-607, 12-713, 14-1879, 21-631, 21-808, 63-680, 221-491, 221-897, 221-986, 221-1051, 248-555, 292-532, 305-943, 316-978, 322-867, 338-994, 353-770, 376-576, 376-897, 378-978, 395-816, 399-978, 399-1147, 406-820, 442-801, 444-1207, 465-1105, 490-912, 495-992, 521-1166, 521-1179, 534-1130, 561-1018, 561-1248, 609-778, 609-1224, 611-1269, 660-1122, 661-1269, 753-1443, 755-1553, 766-1315, 780-1416, 907-1518, 930-1590, 940-1487, 951-1616, 964-1506, 1005-1553, 1012-1875, 1014-1532, 1017-1875, 1021-1637, 1027-1847, 1033-1875, 1046-1875, 1052-1875, 1054-1660, 1074-1801, 1096-1711, 1102-1796, 1109-1875, 1122-1622, 1123-1875, 1134-1716, 1153-1582, 1154-1843, 1155-1823, 1167-1733, 1176-1875, 1190-1801, 1196-1763, 1206-1530, 1216-1875, 1222-1432, 1233-1528, 1233-1835, 1233-1853, 1233-1861, 1233-1867, 1233-1880, 1273-1500, 1273-1523, 1273-1741, 1273-1785, 1287-1818, 1296-1850, 1306-1883, 1326-1750, 1331-1865, 1363-1868, 1389-1633, 1395-1863, 1407-1751, 1421-1751, 1424-1684, 1522-1802, 1554-1878, 1559-1878, 1560-1802, 1560-1883, 1563-1883, 1583-1872, 1596-1851, 1624-1770, 1626-1751, 1631-1824, 1717-1883, 1721-1880, 1766-1878

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
36/7500025CB1/ 2746	1-752, 477-1069, 556-1118, 559-1118, 575-1118, 610-1118, 629-1118, 650-1118, 674-1118, 675-1118, 683-1118, 739-1118, 766-1118, 866-1118, 913-1144, 913-1217, 913-1233, 913-1299, 913-1416, 913-1476, 913-1479, 913-1482, 913-1484, 913-1517, 1004-1563, 1111-1305, 1170-1353, 1174-1450, 1263-1731, 1291-1878, 1292-1927, 1331-2013, 1334-2203, 1336-1635, 1346-1882, 1348-2039, 1350-2145, 1355-1834, 1374-2021, 1396-1667, 1404-1966, 1415-2025, 1416-2102, 1429-1779, 1431-1779, 1437-2142, 1443-2020, 1469-2162, 1480-1587, 1489-2126, 1493-2221, 1523-2313, 1529-1857, 1546-2250, 1555-2026, 1559-2251, 1563-1812, 1592-2294, 1630-2379, 1642-1794, 1643-2474, 1683-2474, 1689-2474, 1699-2474, 1701-2470, 1729-2474, 1735-2474, 1743-2474, 1762-2474, 1770-2474, 1786-2474, 1791-2004, 1796-2474, 1810-2474, 1824-2026, 1833-2474, 1842-2474, 1914-2014, 1914-2474, 1921-2474, 1932-2474, 1957-2474, 2023-2474, 2034-2447, 2046-2474, 2067-2608, 2067-2609, 2073-2374, 2073-2375, 2075-2474, 2076-2562, 2092-2703, 2098-2474, 2110-2576, 2110-2657, 2115-2611, 2147-2343, 2160-2531, 2165-2663, 2165-2710, 2169-2746, 2256-2379, 2265-2662, 2310-2726, 2332-2702
37/7502736CB1/ 2868	1-615, 1-716, 1-719, 1-745, 1-787, 1-828, 1-881, 1-936, 4-750, 6-761, 8-759, 130-783, 144-913, 144-916, 144-921, 148-921, 149-920, 172-920, 177-921, 179-869, 225-999, 383-1065, 406-1071, 409-1054, 452-1071, 452-1076, 457-1254, 467-1126, 467-1336, 470-1180, 471-1130, 471-1210, 477-1327, 530-1364, 533-1164, 533-1467, 534-1300, 537-1335, 554-1239, 561-1234, 562-1328, 570-1344, 580-1526, 585-1277, 593-1317, 602-1239, 627-1569, 633-1308, 635-1320, 667-1292, 709-1482, 721-1316, 725-1433, 732-1460, 733-1443, 733-1522, 758-1427, 758-1599, 784-1529, 794-1522, 801-1392, 812-1677, 813-1441, 814-1511, 818-1539, 818-1546, 818-1559, 822-1556, 822-1562, 831-1444, 838-1563, 839-1444, 871-1656, 871-1669, 889-1782, 892-1729, 896-1687, 897-1599, 899-1731, 905-1747, 910-1753, 917-1676, 942-1531, 954-1676, 971-1791, 1001-1691, 1007-1795, 1007-1799, 1024-1799, 1028-1910, 1029-1806, 1035-1791, 1036-1828, 1046-1937, 1049-1648, 1054-1647, 1061-1953, 1082-1784, 1086-1737, 1094-1959, 1109-2005, 1110-1763, 1114-1844, 1121-1776, 1137-1957, 1148-1914, 1151-2099, 1155-1833, 1161-1792, 1178-1812, 1185-1875, 1195-1932, 1199-1957, 1224-2124, 1229-1993, 1230-2070,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1232-1992, 1232-2061, 1235-1840, 1254-1889, 1280-1854, 1285-2061, 1296-2165, 1298-1975, 1309-2133, 1310-2001, 1312-2107, 1312-2271, 1316-2200, 1321-1947, 1321-2211, 1336-2020, 1340-2196, 1347-2138, 1351-2109, 1352-2069, 1377-1987, 1378-2064, 1379-1941, 1384-1950, 1384-1965, 1399-2104, 1405-1982, 1423-2364, 1427-2009, 1431-2124, 1442-1937, 1451-2088, 1455-2183, 1478-1939, 1485-2150, 1497-2151, 1539-2085, 1549-2436, 1555-2089, 1577-2092, 1582-2096, 1594-2434, 1611-1988, 1619-2011, 1630-1982, 1667-2365, 1727-2436, 1777-2107, 1812-2436, 1883-2436, 1904-2436, 1981-2436, 2008-2436, 2054-2665, 2074-2785, 2127-2625, 2127-2672, 2131-2708, 2143-2708, 2227-2624, 2240-2731, 2246-2438, 2246-2837, 2306-2759, 2323-2868
38/7503570CB1/ 1906	1-257, 1-1847, 19-506, 21-161, 62-598, 63-693, 63-701, 63-744, 63-765, 63-811, 63-836, 63-837, 63-868, 63-870, 63-881, 63-914, 66-460, 66-543, 66-559, 66-576, 66-693, 66-744, 70-591, 70-592, 76-194, 76-211, 76-267, 76-285, 76-310, 76-313, 76-318, 76-321, 76-324, 76-328, 76-330, 76-341, 76-350, 76-351, 76-357, 76-369, 76-377, 76-698, 77-287, 77-300, 77-310, 77-360, 78-288, 78-299, 78-301, 78-318, 78-338, 78-374, 78-379, 78-382, 78-700, 79-239, 79-300, 79-314, 79-317, 79-339, 79-348, 79-357, 79-396, 79-401, 79-508, 80-219, 80-379, 81-314, 81-341, 81-360, 81-679, 82-433, 83-288, 83-327, 83-376, 84-267, 84-276, 84-286, 84-319, 84-337, 84-388, 85-288, 85-323, 85-324, 85-327, 85-331, 85-343, 85-344, 85-346, 85-351, 85-358, 85-380, 85-619, 85-641, 85-669, 86-223, 86-334, 86-346, 86-376, 86-378, 86-380, 86-383, 86-384, 86-415, 86-416, 86-497, 87-292, 87-327, 87-374, 87-513, 87-565, 87-626, 87-741, 87-953, 88-894, 89-381, 89-698, 90-354, 90-788, 92-320, 92-410, 92-786, 93-269, 93-686, 93-787, 101-397, 101-478, 102-327, 103-755, 104-516, 106-365, 107-391, 107-574, 109-356, 109-358, 109-404, 109-532,
	111-274, 111-278, 111-332, 111-369, 111-374, 111-406, 111-410, 111-453, 111-616, 111-633, 111-766, 112-391, 113-312, 113-349, 113-354, 113-380, 115-292, 115-721, 116-346, 116-385, 116-511, 116-659, 116-788, 117-752, 122-674, 122-702, 123-686, 123-888, 128-337, 130-512, 133-518, 133-592, 164-515, 164-718, 167-559, 179-868, 179-923, 186-296, 186-342, 186-358, 186-689, 186-770, 186-791, 207-431, 207-443, 207-516, 207-527, 216-511, 217-798, 227-464, 232-461, 239-1048, 241-682, 250-512, 266-484, 281-481, 284-499, 292-972, 300-564, 305-559, 307-919, 319-608, 320-591, 322-693, 322-888, 326-596, 332-577, 335-1029, 341-601, 341-1115, 343-559, 349-992, 354-1147, 363-992, 371-976, 375-825, 388-695, 393-752, 393-1080, 394-972, 395-885, 398-673, 433-706, 448-1219, 460-949, 461-1062, 465-740, 474-1191, 478-778, 487-732, 488-741, 496-763, 509-756, 519-856, 520-1029, 525-1126, 530-757, 531-761, 531-1039, 532-1026, 536-1290, 540-1238, 568-848, 571-825, 579-1267, 582-859, 586-1323, 588-1167, 594-878, 599-852, 599-864, 599-870, 600-837, 620-882, 623-806, 623-1229, 626-849, 634-860, 650-1103, 658-880, 663-956, 665-903, 668-928, 668-947, 679-992, 680-932, 680-934, 680-966, 684-914, 685-968,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	700-926, 703-1185, 704-963, 704-998, 704-1319, 705-956, 708-988, 708-1021, 723-990, 723-1027, 728-1229, 729-984, 735-975, 736-1273, 737-913, 740-1127, 742-995, 745-1009, 752-1026, 752-1053, 752-1068, 755-1059, 766-1298, 767-949, 767-1015, 773-1029, 776-1284, 776-1322, 783-911, 783-1054, 786-1161, 789-1237, 792-1048, 792-1241, 794-1315, 795-1048, 796-1056, 798-1018, 798-1066, 799-1054, 799-1178, 800-1034, 800-1153, 805-1149, 806-1306, 816-978, 816-992, 816-1063, 827-1107, 835-1127, 841-1323, 863-1323, 866-1143, 873-1147, 878-1153, 879-1185, 880-1323, 882-1172, 889-1323, 892-1137, 892-1158, 895-1140, 898-1166, 904-1144, 904-1323, 911-1158, 911-1178, 916-1174, 926-1194, 928-1140, 928-1172, 941-1193, 942-1174, 945-1208, 945-1209, 947-1042, 949-1212, 956-1239, 957-1165, 959-1323, 960-1224, 960-1233, 960-1277, 960-1323, 961-1198, 961-1222, 961-1322, 962-1186, 963-1156, 963-1216, 964-1213, 966-1213, 969-1323, 972-1270, 976-1536, 976-1603, 980-1231, 980-1272, 980-1306, 981-1263, 981-1265, 981-1308, 981-1323, 982-1323, 991-1240, 992-1114, 997-1318, 1002-1241, 1002-1323, 1004-1248, 1006-1239, 1007-1323, 1009-1251, 1010-1323, 1012-1303, 1022-1323, 1023-1272,
	1027-1262, 1030-1291, 1034-1297, 1036-1299, 1043-1323, 1045-1192, 1047-1294, 1047-1301, 1047-1323, 1048-1281, 1049-1286, 1050-1274, 1051-1323, 1054-1264, 1054-1312, 1064-1193, 1064-1253, 1064-1322, 1064-1323, 1066-1321, 1084-1323, 1092-1299, 1095-1323, 1096-1323, 1101-1323, 1112-1304, 1116-1322, 1134-1323, 1145-1323, 1151-1278, 1167-1323, 1172-1323, 1195-1286, 1201-1274, 1201-1323, 1309-1337, 1335-1480, 1335-1490, 1335-1536, 1335-1553, 1335-1558, 1335-1559, 1335-1570, 1335-1591, 1335-1623, 1335-1762, 1335-1782, 1335-1841, 1335-1855, 1335-1869, 1335-1899, 1337-1412, 1337-1602, 1338-1570, 1338-1590, 1338-1627, 1338-1762, 1339-1597, 1340-1649, 1342-1848, 1345-1590, 1345-1597, 1346-1577, 1347-1623, 1347-1628, 1349-1583, 1349-1645, 1350-1906, 1352-1901, 1353-1561, 1353-1858, 1354-1830, 1363-1524, 1366-1892, 1371-1641, 1385-1791, 1386-1667, 1392-1626, 1394-1859, 1395-1669, 1395-1859, 1395-1874, 1396-1656, 1396-1675, 1402-1665, 1403-1671, 1404-1556, 1405-1783, 1406-1661, 1406-1673, 1411-1888, 1413-1901, 1422-1745, 1422-1901, 1427-1883, 1430-1626, 1433-1678, 1435-1882, 1435-1893, 1439-1858, 1439-1900, 1441-1707, 1441-1855, 1441-1862,
	1442-1676, 1443-1864, 1445-1862, 1447-1762, 1451-1723, 1453-1897, 1454-1727, 1455-1858, 1455-1859, 1457-1859, 1462-1888, 1469-1762, 1475-1762, 1479-1861, 1480-1694, 1480-1767, 1480-1853, 1480-1878, 1482-1762, 1483-1858, 1487-1762, 1490-1762, 1490-1771, 1492-1901, 1498-1856, 1499-1864, 1501-1535, 1502-1858, 1505-1762, 1509-1840, 1514-1744, 1520-1800, 1520-1891, 1525-1830, 1528-1750, 1528-1756, 1542-1858, 1546-1835, 1576-1829, 1578-1788, 1583-1861, 1592-1762, 1593-1762, 1594-1835, 1615-1762, 1645-1859, 1647-1762, 1649-1885, 1677-1901, 1696-1836

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
39/7504008CB1/ 2506	1-646, 4-684, 4-805, 6-692, 12-838, 21-242, 21-299, 21-553, 21-566, 21-656, 21-763, 21-767, 21-795, 21-805, 21-840, 21-2480, 22-744, 24-583, 45-321, 124-640, 151-691, 196-807, 276-639, 343-840, 415-823, 465-639, 526-740, 526-741, 551-1334, 611-1400, 638-1560, 715-1311, 840-1310, 840-1462, 870-1096, 873-1696, 889-1374, 898-1565, 920-1880, 944-1563, 948-1584, 958-1588, 960-1188, 960-1435, 972-1830, 977-1310, 982-1531, 999-1909, 1002-1519, 1033-1349, 1033-1626, 1033-1678, 1035-1586, 1037-1173, 1045-1529, 1049-1494, 1064-1332, 1065-1761, 1066-1612, 1084-1735, 1096-1753, 1103-1241, 1110-1724, 1115-1688, 1127-1888, 1148-1678, 1151-1382, 1152-1668, 1152-1710, 1159-1432, 1159-1683, 1182-1758, 1195-1736, 1195-1818, 1205-1707, 1220-1461, 1227-1783, 1228-1787, 1228-1830, 1228-1877, 1237-1523, 1237-1877, 1237-1901, 1237-2084, 1243-1467, 1245-1765, 1254-1673, 1264-1498, 1268-1795, 1272-1556, 1277-1790, 1310-1560, 1325-1883, 1342-2008, 1346-2008, 1352-1692, 1352-1693, 1366-1943, 1369-1485, 1370-1822, 1384-1985, 1385-1576, 1391-2191, 1408-2018, 1408-2041, 1409-1987, 1413-1698, 1424-1959, 1426-1659, 1427-1687, 1431-1976, 1439-2004, 1448-2087, 1454-1706,
	1460-2004, 1465-1973, 1467-1744, 1468-1778, 1469-2009, 1473-1993, 1473-2142, 1477-1777, 1480-2060, 1498-2028, 1502-2056, 1510-2040, 1523-1851, 1524-1703, 1524-2048, 1531-1721, 1538-1784, 1542-2171, 1574-1828, 1574-1984, 1596-2175, 1599-2191, 1601-1794, 1604-2478, 1615-2029, 1615-2159, 1623-2156, 1636-1856, 1643-2184, 1648-1846, 1648-2123, 1648-2321, 1651-2161, 1654-1911, 1660-1900, 1668-1932, 1670-2478, 1672-2276, 1674-2478, 1675-2358, 1675-2458, 1677-2478, 1684-2478, 1686-2478, 1710-2478, 1711-2131, 1711-2449, 1716-2478, 1726-2334, 1727-2480, 1731-1977, 1731-2222, 1731-2458, 1736-2012, 1741-2380, 1753-2480, 1773-2324, 1777-2478, 1781-2054, 1781-2203, 1781-2207, 1781-2211, 1784-2079, 1786-1996, 1789-2373, 1791-2478, 1800-2064, 1806-2403, 1810-2076, 1811-2408, 1818-2406, 1834-2052, 1834-2102, 1837-2454, 1837-2462, 1840-2343, 1841-1974, 1847-2478, 1851-2134, 1851-2145, 1853-2332, 1858-2491, 1859-2281, 1862-2416, 1865-2492, 1866-2091, 1872-2436, 1872-2491, 1880-2463, 1880-2491, 1882-2344, 1892-2335, 1892-2472, 1892-2477, 1895-2108, 1899-2469, 1905-2450, 1906-2506, 1917-2495, 1923-2461, 1929-2464, 1929-2492, 1935-2439, 1939-2500,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1946-2404, 1948-2498, 1951-2133, 1956-2284, 1956-2471, 1959-2466, 1981-2500, 1984-2314, 1986-2493, 1988-2131, 1990-2144, 1990-2449, 1993-2292, 2011-2337, 2011-2480, 2012-2497, 2016-2250, 2018-2480, 2038-2273, 2043-2500, 2047-2478, 2054-2322, 2055-2472, 2055-2480, 2055-2481, 2066-2478, 2068-2419, 2068-2497, 2072-2497, 2074-2478, 2079-2499, 2082-2478, 2087-2478, 2095-2470, 2108-2477, 2109-2478, 2120-2470, 2128-2427, 2137-2485, 2140-2480, 2155-2478, 2164-2478, 2180-2383, 2182-2467, 2187-2479, 2189-2339, 2203-2483, 2205-2484, 2206-2478, 2220-2498, 2221-2499, 2225-2478, 2229-2480, 2240-2476, 2249-2478, 2252-2479, 2259-2478, 2268-2478, 2275-2392, 2283-2476, 2297-2479, 2325-2446, 2327-2476
40/7503559CB1/ 3836	1-202, 1-3836, 64-721, 170-331, 170-712, 170-746, 170-826, 580-1161, 777-1229, 872-1387, 873-1130, 962-1221, 1210-1797, 1213-1964, 1435-1710, 1535-1793, 1684-2178, 1710-2577, 1733-2214, 1740-2206, 1780-2054, 1800-2095, 1832-2577, 1874-2577, 1899-2577, 1918-2577, 1944-2577, 1968-2577, 2029-2577, 2152-2577, 2301-2660, 2329-2487, 2330-2852, 2440-3039, 2589-3210, 2619-3480, 2790-3490, 2792-3083, 2870-3374, 2947-3208, 2961-3405, 2983-3788, 3032-3276, 3137-3404, 3193-3529, 3353-3531, 3366-3672, 3366-3681, 3497-3787, 3524-3833, 3532-3763
41/6243872CB1/ 2240	1-651, 8-746, 23-651, 196-951, 196-952, 196-960, 196-1116, 429-1277, 454-1277, 585-1277, 1074-1751, 1652-2163, 1652-2177, 1652-2180, 1652-2191, 1652-2238, 1652-2239, 1652-2240, 1658-2174
42/90011608CB1/ 2807	1-784, 719-1600, 738-1509, 1330-2220, 1388-2204, 2169-2807
43/90024583CB1/ 3201	1-857, 1-3192, 669-1398, 978-1896, 1321-2123, 1849-2528, 2368-3201

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
44/90113658CB1/ 3688	1-669, 1-1527, 397-425, 512-1174, 888-1174, 922-1119, 922-1127, 1087-1594, 1088-1347, 1098-1850, 1241-2121, 1245-1943, 1252-1821, 1252-1861, 1252-1872, 1252-1948, 1252-1950, 1252-1978, 1252-1999, 1252-2015, 1252-2025, 1252-2060, 1252-2064, 1252-2075, 1252-2076, 1252-2081, 1252-2121, 1252-2123, 1252-2125, 1252-2137, 1252-2141, 1252-2159, 1252-2168, 1252-2171, 1252-2177, 1252-2233, 1253-1948, 1253-1991, 1253-2027, 1253-2058, 1253-2250, 1255-1915, 1347-1610, 1361-1556, 1384-2129, 1486-1759, 1489-1762, 1574-1896, 1574-1967, 1574-2030, 1574-2122, 1574-2154, 1574-2168, 1574-2187, 1660-1917, 1695-1989, 1699-1870, 1702-2193, 1766-2118, 1789-2051, 1789-2208, 1805-2748, 1809-2748, 1824-3673, 1830-2748, 1835-2748, 1841-1976, 1859-2744, 1869-2747, 1871-2390, 1878-2748, 1879-2067, 1879-2540, 1880-2748, 1932-2583, 1932-2753, 1933-2191, 1933-2499, 1933-2748, 1961-2748, 1968-2748, 1973-2748, 1999-2748, 2030-2748, 2082-2853, 2100-2134, 2100-2169, 2203-2476, 2203-2656, 2234-2851, 2266-2748, 2272-2575, 2272-2748, 2279-2983, 2360-2823, 2389-3176, 2393-2731, 2439-2899, 2442-2898, 2461-3112, 2532-2748, 2551-2899, 2597-3176, 2653-2896, 2657-2945, 2657-3078,
	2757-3688, 2800-3461, 2806-3189, 2921-3208, 2979-3208, 3002-3660, 3049-3323, 3049-3336
45/3942766CB1/ 2402	1-119, 1-1900, 102-557, 115-820, 138-365, 138-670, 245-518, 287-350, 354-936, 380-742, 380-769, 706-792, 706-1288, 829-1409, 1034-1521, 1244-1359, 1267-1856, 1433-1607, 1504-1620, 1512-2003, 1567-1820, 1567-2402, 1717-2042, 1745-1898
46/7501987CB1/ 2410	1-351, 1-2383, 267-538, 267-556, 267-591, 267-733, 267-745, 267-786, 267-790, 267-801, 267-805, 267-806, 267-843, 267-844, 309-825, 394-978, 404-1062, 404-1128, 404-1159, 404-1183, 404-1194, 404-1196, 404-1197, 404-1286, 412-1024, 419-1194, 443-1089, 513-1120, 517-989, 544-700, 551-1100, 608-1167, 616-1158, 622-1046, 643-888, 643-1282, 658-1214, 662-1142, 668-1304, 682-971, 698-1284, 744-1111, 745-1389, 747-1218, 750-1420, 751-1008, 786-1402, 793-1445, 801-903, 812-1397, 832-1446, 860-1090, 869-1379, 880-1022, 908-1158, 912-1549, 920-1554, 923-1515, 932-1545, 951-1503, 958-1567, 962-1500, 1024-1588, 1052-1291, 1108-1519, 1108-1643, 1109-1393, 1109-1499, 1113-1650, 1130-1503, 1163-1742, 1185-1721, 1192-1614, 1193-1458, 1193-1737, 1195-1832, 1196-1754, 1204-1837, 1212-1786, 1219-1648, 1224-1720, 1241-1956, 1244-1825, 1249-1874, 1254-1896, 1255-1865, 1257-1692, 1267-1668, 1283-1792, 1296-1559, 1296-1565, 1321-1900, 1324-1619, 1334-1908, 1348-1868, 1357-1976, 1364-1521, 1374-1803, 1384-1721, 1388-2036, 1397-2287, 1401-2164, 1407-1612, 1407-1648, 1407-2038, 1411-1975, 1415-1999, 1418-1574, 1425-1694, 1460-2126, 1465-1729, 1468-2170, 1471-2164,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1475-1751, 1475-1935, 1475-2009, 1475-2027, 1506-1794, 1512-2021, 1523-2406, 1537-1792, 1542-2161, 1556-1864, 1558-2230, 1559-2011, 1559-2410, 1578-2392, 1580-1951, 1591-2059, 1597-2106, 1597-2410, 1616-1939, 1617-2205, 1623-2360, 1627-2410, 1631-1903, 1634-2122, 1643-2242, 1651-2256, 1655-2258, 1660-2410, 1699-2160, 1701-2364, 1704-2351, 1706-2272, 1714-2383, 1715-2033, 1715-2133, 1715-2210, 1715-2284, 1715-2288, 1715-2363, 1716-2258, 1753-2283, 1785-2107, 1785-2207, 1785-2372, 1789-2373, 1800-2080, 1805-2068, 1805-2267, 1812-2317, 1819-2311, 1828-2362, 1872-2383, 1898-2146, 1908-2164, 1911-2157, 1915-2336, 1915-2338, 1916-2406, 1933-2392, 1984-2252, 1994-2344, 2008-2316, 2039-2334, 2078-2360, 2100-2383, 2149-2330, 2157-2317, 2299-2350
47/7503223CB1/ 968	1-244, 1-584, 1-772, 1-797, 1-831, 1-968, 2-796, 4-888, 10-888, 26-513, 31-888, 117-888, 284-537, 547-968, 548-732, 548-756, 548-776, 548-809, 548-827, 548-960, 548-968, 557-968, 575-968, 581-850, 590-952, 590-968, 618-968, 653-952, 671-937, 673-968, 715-959, 734-952, 758-952, 765-952, 768-952, 818-968, 819-955
48/7503566CB1/ 2267	1-245, 1-531, 1-551, 1-576, 1-2267, 5-814, 7-544, 7-588, 7-683, 7-688, 7-763, 7-764, 7-772, 8-611, 16-596, 18-681, 27-306, 28-283, 45-622, 197-678, 288-426, 335-573, 340-923, 367-1029, 388-1032, 388-1101, 414-1332, 419-688, 420-551, 421-937, 444-1112, 491-960, 491-1076, 544-1031, 598-1179, 603-971, 603-974, 603-1171, 640-1166, 741-1261, 756-929, 756-985, 756-1230, 756-1262, 756-1269, 756-1274, 756-1282, 756-1308, 782-1323, 832-1364, 853-1133, 859-1418, 969-1502, 1022-1602, 1047-1611, 1050-1404, 1051-1274, 1177-1642, 1316-1640, 1316-1774, 1316-1853, 1316-1859, 1316-1875, 1317-1601, 1317-1751, 1317-1800, 1317-1830, 1317-1844, 1317-1851, 1317-1852, 1317-1860, 1317-1942, 1317-1964, 1317-2191, 1318-1981, 1376-1622, 1376-1897, 1397-2251, 1417-1751, 1417-1853, 1417-1948, 1427-1675, 1430-2254, 1458-1959, 1460-1732, 1466-2251, 1471-1678, 1475-1706, 1475-2034, 1545-2254, 1553-2251, 1555-1794, 1561-2254, 1563-1804, 1564-2254, 1567-2254, 1570-2254, 1579-2034, 1580-2034, 1593-1815, 1608-2034, 1610-1876, 1617-2133, 1631-2253, 1659-2181, 1672-2200, 1711-1927, 1716-1942, 1716-1946, 1721-1951, 1727-2129, 1733-1945, 1756-1960, 1759-2180, 1761-2181, 1768-2159, 1807-2251,
	1812-2051, 1816-2256, 1824-2265, 1843-2119, 1849-2075, 1854-1932, 1854-1947, 1864-2088, 1933-2180, 1954-2205, 1994-2223, 1998-2235
49/7505122CB1/ 319	1-217, 1-319, 15-216, 39-148, 39-200, 39-211, 39-214, 39-216, 39-223, 40-161, 40-223, 40-294, 40-298, 41-223, 45-158, 46-173, 49-147, 58-223, 63-223, 90-208, 221-306

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
50/7511620CB1/ 2510	1-222, 1-279, 1-328, 1-340, 1-373, 1-375, 1-379, 1-394, 1-412, 1-415, 1-416, 1-423, 1-432, 1-440, 1-458, 1-462, 1-2510, 4-339, 25-301, 256-465, 276-507, 276-812, 431-465, 464-662, 464-926, 465-812, 465-1030, 465-1367, 467-661, 469-1121, 474-1260, 479-1287, 481-1115, 483-990, 534-1443, 549-1375, 549-1409, 557-1329, 559-1238, 560-1258, 561-1308, 573-1109, 579-1332, 588-1389, 592-1194, 596-1436, 621-1067, 622-1333, 628-1022, 630-1378, 646-1439, 647-1275, 651-1420, 656-1425, 657-1488, 658-1369, 659-1330, 661-1107, 661-1290, 673-1362, 675-1360, 690-1330, 711-1460, 713-1178, 717-1150, 717-1344, 718-1571, 722-1442, 729-1523, 729-1526, 730-1452, 736-1532, 740-1656, 748-1261, 752-1593, 756-1000, 762-1599, 765-1546, 779-1200, 779-1583, 779-1672, 796-1580, 796-1601, 801-1392, 802-1591, 803-1589, 813-1048, 813-1190, 813-1364, 813-1367, 813-1388, 816-1708, 817-1478, 846-1321, 850-1345, 851-1538, 858-1497, 858-1794, 867-1595, 882-1523, 888-1556, 893-1750, 902-1751, 905-1131, 906-1566, 908-1555, 908-1731, 908-1795, 909-1800, 933-1600, 937-1701, 939-1698, 939-1701,
	940-1783, 955-1912, 965-1892, 967-1851, 968-1757, 977-1795, 979-1598, 983-1619, 986-1403, 989-1678, 992-1754, 993-1623, 994-1788, 995-1402, 1007-1870, 1008-1873, 1012-1345, 1015-1884, 1016-1757, 1016-1795, 1017-1566, 1022-1789, 1023-1717, 1037-1554, 1044-1913, 1049-1830, 1049-1861, 1051-1870, 1053-1636, 1053-1787, 1056-1888, 1056-1892, 1065-2046, 1068-1384, 1068-1661, 1068-1713, 1070-1621, 1072-1208, 1075-1839, 1076-1784, 1080-1564, 1084-1529, 1087-1863, 1089-1788, 1089-1927, 1091-1765, 1091-1891, 1093-1776, 1099-1367, 1100-1796, 1101-1647, 1103-1747, 1118-1782, 1119-1770, 1119-1775, 1119-1873, 1120-1705, 1121-1850, 1128-1783, 1131-1788, 1135-1952, 1135-2012, 1136-1958, 1138-1276, 1145-1759, 1147-1843, 1150-1723, 1152-1865, 1159-2027, 1162-1920, 1183-1713, 1186-1417, 1187-1703, 1187-1745, 1194-1467, 1194-1638, 1196-1763, 1196-1947, 1206-1851, 1213-1869, 1217-1793, 1229-2043, 1230-1771, 1230-1853, 1232-2081, 1233-2062, 1240-1742, 1255-1496, 1262-1818, 1263-1822, 1263-1865, 1263-1909, 1267-2001, 1272-1558, 1272-1838, 1274-1919, 1278-1502, 1279-2065, 1280-1800, 1289-1708, 1293-1915, 1294-1998, 1299-1533, 1299-2063, 1301-1903, 1303-1830,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	1307-1591, 1312-1825, 1314-2137, 1351-2083, 1360-1915, 1371-2030, 1377-2040, 1381-2040, 1387-1727, 1387-1728, 1398-1938, 1401-1975, 1404-1520, 1405-1857, 1414-2072, 1419-2017, 1420-1611, 1420-2209, 1424-2143, 1426-2223, 1438-2065, 1443-2050, 1443-2073, 1444-2019, 1448-1733, 1461-1694, 1466-2008, 1471-1687, 1483-2119, 1484-2210, 1489-1741, 1492-2147, 1500-2005, 1502-1779, 1503-1813, 1504-2041, 1508-2025, 1521-2320, 1533-2060, 1537-2088, 1545-2072, 1559-1738, 1559-2080, 1566-1756, 1573-1819, 1577-2203, 1581-2173, 1585-2421, 1594-1749, 1609-1863, 1623-2466, 1630-2383, 1631-2207, 1634-2223, 1636-1829, 1639-2510, 1646-2331, 1653-1758, 1657-2349, 1666-2510, 1670-2292, 1678-2216, 1678-2354, 1683-1878, 1683-2155, 1683-2353, 1683-2390, 1684-2510, 1686-2510, 1687-2409, 1687-2510, 1688-2510, 1689-1943, 1689-2510, 1692-2424, 1692-2501, 1705-2470, 1705-2510, 1707-2308, 1709-2510, 1710-2367, 1711-2509, 1712-2510, 1719-2510, 1721-2510, 1728-2510, 1730-2510, 1734-2510, 1745-2510, 1746-2163, 1746-2481, 1750-2268, 1751-2510, 1752-2510, 1753-2449, 1761-2366, 1762-2510, 1766-2009, 1766-2267, 1766-2490, 1776-2412, 1778-2510, 1779-2510, 1781-2510,
	1787-2510, 1788-2510, 1792-2510, 1795-2510, 1800-2510, 1808-2356, 1810-2510, 1812-2510, 1815-2510, 1816-2086, 1816-2178, 1816-2235, 1816-2239, 1816-2243, 1816-2402, 1816-2510, 1817-2359, 1818-2062, 1818-2510, 1824-2405, 1824-2510, 1825-2510, 1826-2510, 1831-2510, 1835-2096, 1835-2510, 1841-2435, 1842-2510, 1844-2510, 1845-2108, 1853-2438, 1853-2459, 1858-2510, 1865-2032, 1865-2152, 1865-2471, 1865-2490, 1865-2510, 1866-2084, 1866-2134, 1871-2510, 1872-2375, 1873-2006, 1879-2510, 1882-2510, 1883-2166, 1883-2177, 1884-2510, 1885-2364, 1888-2510, 1889-2510, 1890-2501, 1891-2313, 1894-2448, 1896-2510, 1897-2510, 1898-2123, 1898-2510, 1903-2510, 1904-2468, 1904-2510, 1910-2510, 1912-2495, 1912-2509, 1914-2376, 1923-2510, 1924-2367, 1924-2509, 1926-2510, 1927-2140, 1929-2510, 1930-2510, 1931-2510, 1932-2510, 1937-2482, 1937-2510, 1938-2510, 1944-2484, 1949-2356, 1949-2510, 1960-2510, 1961-2467, 1961-2510, 1964-2483, 1967-2471, 1971-2510, 1974-2510, 1977-2510, 1980-2498, 1981-2510, 1982-2510, 1983-2165, 1983-2510, 1988-2253, 1988-2286, 1990-2437, 1992-2510, 1996-2510, 2000-2510, 2002-2510, 2011-2510, 2013-2510, 2016-2346, 2016-2473,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	2018-2510, 2020-2163, 2022-2176, 2022-2481, 2025-2324, 2029-2510, 2033-2510, 2039-2510, 2040-2510, 2041-2510, 2042-2259, 2043-2368, 2043-2472, 2043-2510, 2044-2510, 2048-2282, 2050-2510, 2059-2510, 2065-2275, 2068-2510, 2070-2305, 2075-2510, 2079-2471, 2079-2510, 2084-2510, 2085-2510, 2086-2354, 2087-2504, 2087-2510, 2094-2510, 2098-2510, 2099-2481, 2100-2510, 2104-2510, 2105-2510, 2106-2439, 2106-2510, 2111-2510, 2114-2510, 2119-2510, 2122-2510, 2123-2510, 2127-2510, 2128-2510, 2131-2476, 2140-2509, 2141-2510, 2144-2510, 2152-2502, 2152-2510, 2160-2459, 2169-2510, 2172-2510, 2177-2510, 2179-2510, 2187-2510, 2189-2406, 2189-2510, 2196-2510, 2212-2415, 2214-2499, 2217-2510, 2219-2510, 2220-2510, 2221-2371, 2235-2510, 2237-2510, 2238-2510, 2252-2510, 2253-2510, 2257-2510, 2261-2510, 2272-2510, 2281-2510, 2284-2510, 2291-2510, 2300-2510, 2315-2510, 2329-2510, 2338-2463, 2338-2510, 2357-2478, 2359-2510, 2375-2510
51/7506995CB1/ 2241	1-351, 1-2214, 267-538, 267-556, 267-591, 267-717, 267-718, 294-1269, 544-700, 719-853, 719-921, 719-1210, 739-989, 743-1380, 751-1385, 754-1346, 763-1376, 782-1334, 789-1398, 793-1331, 821-1750, 855-1419, 862-1751, 883-1122, 884-1750, 895-1751, 896-1751, 928-1751, 938-1350, 939-1474, 940-1224, 940-1330, 944-1481, 946-1751, 949-1750, 961-1334, 961-1750, 966-1750, 986-1751, 988-1750, 994-1573, 1005-1750, 1016-1552, 1023-1445, 1024-1289, 1024-1568, 1024-1750, 1027-1585, 1028-1749, 1031-1663, 1035-1668, 1042-1750, 1043-1617, 1044-1750, 1050-1479, 1055-1551, 1060-1751, 1062-1750, 1065-1750, 1072-1787, 1075-1656, 1080-1705, 1085-1727, 1086-1696, 1088-1523, 1098-1499, 1114-1623, 1115-1745, 1127-1390, 1127-1396, 1138-1750, 1152-1731, 1155-1450, 1165-1739, 1174-1807, 1182-1699, 1195-1352, 1205-1634, 1215-1552, 1219-1867, 1228-2118, 1232-1995, 1238-1443, 1238-1479, 1238-1869, 1242-1806, 1246-1830, 1249-1405, 1256-1525, 1291-1957, 1296-1560, 1299-2001, 1302-1995, 1306-1582, 1306-1766, 1306-1840, 1306-1858, 1337-1625, 1343-1852, 1354-2237, 1368-1623, 1373-1992, 1387-1695, 1389-2061, 1390-1842, 1390-2241, 1409-2223, 1411-1782, 1422-1890, 1428-1937,
	1428-2241, 1447-1770, 1448-2036, 1454-2191, 1458-2241, 1462-1734, 1465-1953, 1474-2073, 1482-2087, 1486-2089, 1491-2241, 1530-1991, 1532-2195, 1535-2182, 1537-2103, 1545-2214, 1546-1864, 1546-1964, 1546-2041, 1546-2115, 1546-2119, 1546-2194, 1547-2089, 1584-2114, 1616-1938, 1616-2038, 1616-2203, 1620-2204, 1631-1911, 1636-1899, 1636-2098, 1643-2148, 1650-2142, 1659-2193, 1703-2214, 1729-1977, 1739-1995, 1742-1988, 1746-2167, 1746-2169, 1747-2237, 1764-2223, 1815-2083, 1825-2175, 1839-2147, 1870-2165, 1909-2191, 1931-2214, 1980-2161, 1988-2148, 2130-2181

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
52/7506996CB1/ 2312	1-353, 1-2285, 269-540, 269-558, 269-593, 269-735, 269-747, 269-788, 269-792, 269-803, 269-807, 269-808, 269-845, 269-846, 295-859, 295-884, 295-970, 295-983, 295-1011, 295-1013, 295-1076, 295-1117, 295-1144, 295-1155, 295-1195, 295-1199, 296-1078, 311-827, 396-980, 406-1196, 414-1026, 445-1091, 515-1122, 519-991, 546-702, 553-1102, 610-1169, 618-1160, 624-1048, 645-890, 660-1197, 664-1144, 684-973, 746-1113, 753-1010, 862-1092, 882-1024, 910-1160, 971-1822, 1198-1461, 1198-1467, 1209-1821, 1223-1802, 1226-1521, 1236-1810, 1245-1878, 1253-1770, 1266-1423, 1276-1705, 1286-1623, 1290-1938, 1299-2189, 1303-2066, 1309-1514, 1309-1550, 1309-1940, 1313-1877, 1317-1901, 1327-1596, 1362-2028, 1367-1631, 1370-2072, 1373-2066, 1377-1653, 1377-1837, 1377-1911, 1377-1929, 1391-2024, 1408-1696, 1414-1923, 1425-2308, 1439-1694, 1444-2063, 1458-1766, 1460-2132, 1461-1913, 1461-2312, 1480-2294, 1482-1853, 1493-1961, 1499-2008, 1499-2312, 1518-1841, 1519-2107, 1525-2262, 1529-2312, 1533-1805, 1536-2024, 1545-2144, 1553-2158, 1557-2160, 1562-2312, 1601-2062, 1603-2266, 1606-2253, 1608-2174, 1616-2285, 1617-1935, 1617-2035, 1617-2112, 1617-2186, 1617-2190,
	1617-2265, 1618-2160, 1655-2185, 1687-2009, 1687-2274, 1691-2275, 1702-1982, 1707-1970, 1707-2169, 1714-2219, 1721-2213, 1730-2264, 1774-2285, 1800-2048, 1810-2066, 1813-2059, 1817-2238, 1817-2240, 1818-2308, 1835-2294, 1886-2154, 1896-2246, 1910-2218, 1941-2236, 1980-2262, 2002-2285, 2059-2219

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
27	1853191CB1	PROSNOT15
28	7497369CB1	KIDEUNE02
29	1700438CB1	BLADTUT05
30	535939CB1	HEARFET01
31	55118067CB1	LIVRTUT13
33	7500819CB1	CORPNOT02
34	7503413CB1	LPARNOT02
35	7500007CB1	BRAITUT02
36	7500025CB1	COLCTUT02
37	7502736CB1	BRANDIN01
38	7503570CB1	CARGDIT01
39	7504008CB1	THP1TXT03
40	7503559CB1	MCLDTXN03
41	6243872CB1	TESTNOT17
44	90113658CB1	PROSTUT04
45	3942766CB1	FIBRTXS07
46	7501987CB1	BRAENOT04
47	7503223CB1	PROSTUT10
48	7503566CB1	BSTMNON02
49	7505122CB1	SINTNOT13
50	7511620CB1	COLNNOT23
51	7506995CB1	BRAENOT04
52	7506996CB1	BRAENOT04

Table 6

Library	Vector	Library Description
	NONE	NONE
BLADTUT05	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma on the anterior wall of the bladder. Patient history included lung neoplasm and tobacco abuse in remission. Family history included malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer.
BRAENOT04	pINCY	Library was constructed using RNA isolated from inferior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.
BRAITUT02	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
BRANDIN01	pINCY	This normalized pineal gland tissue library was constructed from .4 million independent clones from a pineal gland tissue library from two different donors. Starting RNA was made from pooled pineal gland tissue removed from two Caucasian females: a 68-year-old (donor A) who died from congestive heart failure and a 79-year-old (donor B) who died from pneumonia. Neuropathology for donor A indicated mild to moderate Alzheimer disease, atherosclerosis, and multiple infarctions. Neuropathology for donor B indicated severe Alzheimer disease, arteriolosclerosis, cerebral amyloid angiopathy and multiple infarctions. There were diffuse and neuritic amyloid plaques and neurofibrillary tangles throughout the brain sections examined in both donors. Patient history included diabetes mellitus, rheumatoid arthritis, hyperthyroidism, amyloid heart disease, and dementia in donor A; and pseudophakia, gastritis with bleeding, glaucoma, peripheral vascular disease, COPD, delayed onset tonic/clonic seizures, and transient ischemic attack in donor B. The library was normalized in one round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo e
BSTMNON02	PSPORT1	This normalized brain stem library was constructed from 2.84 million independent clones from a brain stem library. Starting RNA was made from the brain stem tissue of a 72-year-old Caucasian male who died from myocardial infarction. Patient history included coronary artery disease, insulin-dependent diabetes mellitus, and arthritis. Normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9228).
CARGDIT01	pINCY	Library was constructed using RNA isolated from diseased cartilage tissue. Patient history included osteoarthritis.

Table 6

Library	Vector	Library Description
COLCTUT02	pINCY	Library was constructed using RNA isolated from colon tumor tissue removed from the cecum of a 30-year-old Caucasian female during partial colectomy, open liver biopsy, incidental appendectomy, and permanent colostomy. Pathology indicated carcinoid tumor (grade 1 neuroendocrine carcinoma) arising in the terminal ileum, forming a mass in the right colon. Patient history included chronic sinus infections and endometriosis. Family history included hyperlipidemia, anxiety, upper lobe lung cancer, stomach cancer, liver cancer, and cirrhosis.
COLNNOT23	pINCY	Library was constructed using RNA isolated from diseased colon tissue removed from a 16-year-old Caucasian male during a total colectomy with abdominal/perineal resection. Pathology indicated gastritis and pancolonitis consistent with the acute phase of ulcerative colitis. Inflammation was more severe in the transverse colon, with inflammation confined to the mucosa. There was only mild involvement of the ascending and sigmoid colon, and no significant involvement of the cecum, rectum, or terminal ileum. Family history included irritable bowel syndrome.
CORPNOT02	pINCY	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
FIBRTXS07	pINCY	This subtracted library was constructed using 1.3 million clones from a dermal fibroblast library and was subjected to two rounds of subtraction hybridization with 2.8 million clones from an untreated dermal fibroblast tissue library. The starting library for subtraction was constructed using RNA isolated from treated dermal fibroblast tissue removed from the breast of a 31-year-old Caucasian female. The cells were treated with 9CIS retinoic acid. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated dermal fibroblast tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.
HEARFET01	pINCY	Library was constructed using RNA isolated from heart tissue removed from a Hispanic male fetus, who died at 18 weeks' gestation.
KIDEUNE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from an untreated transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue (Invitrogen). The cells were transformed with adenovirus 5 DNA.

Table 6

Library	Vector	Library Description
LIVRTUT13	pINCY	Library was constructed using RNA isolated from liver tumor tissue removed from a 62-year-old Caucasian female during partial hepatectomy and exploratory laparotomy. Pathology indicated metastatic intermediate grade neuroendocrine carcinoma, consistent with islet cell tumor, forming nodules ranging in size, in the lateral and medial left liver lobe. The pancreas showed fibrosis, chronic inflammation and fat necrosis consistent with pseudocyst. The gall bladder showed mild chronic cholecystitis. Patient history included malignant neoplasm of the pancreas tail, pulmonary embolism, hyperlipidemia, thrombophlebitis, joint pain in multiple joints, type II diabetes, benign hypertension, and cerebrovascular disease. Family history included pancreas cancer, secondary liver cancer, benign hypertension, and hyperlipidemia.
LPARNOT02	pINCY	Library was constructed using RNA isolated from tissue obtained from the left parotid (salivary) gland of a 70-year-old male with parotid cancer.
MCLDTXN03	pINCY	This normalized dendritic cell library was constructed from one million independent clones from a pool of two derived dendritic cell libraries. Starting libraries were constructed using RNA isolated from untreated and treated derived dendritic cells from umbilical cord blood CD34+ precursor cells removed from a male. The cells were derived with granulocyte/macrophage colony stimulating factor (GM-CSF), tumor necrosis factor alpha (TNF alpha), and stem cell factor (SCF). The GM-CSF was added at time 0 at 100 ng/ml, the TNF alpha was added at time 0 at 2.5 ng/ml, and the SCF was added at time 0 at 25 ng/ml. Incubation time was 13 days. The treated cells were then exposed to phorbol myristate acetate (PMA), and Ionomycin. The PMA and Ionomycin were added at 13 days for five hours. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PROSNOT15	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 66-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 2+3). The patient presented with elevated prostate specific antigen (PSA). Family history included prostate cancer, secondary bone cancer, and benign hypertension.
PROSTUT04	PSPORT1	Library was constructed using RNA isolated from prostate tumor tissue removed from a 57-year-old Caucasian male during radical prostatectomy, removal of both testes and excision of regional lymph nodes. Pathology indicated adenocarcinoma (Gleason grade 3+3). Patient history included a benign neoplasm of the large bowel and type I diabetes. Family history included a malignant neoplasm of the prostate and type I diabetes.

Table 6

Library	Vector	Library Description
PROSTUT10	pINCY	Library was constructed using RNA isolated from prostatic tumor tissue removed from a 66-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated an adenocarcinoma (Gleason grade 2+3). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Family history included prostate cancer and secondary bone cancer.
SINTNOT13	pINCY	Library was constructed using RNA isolated from ileum tissue obtained from a 25-year-old Asian female during a partial colectomy and temporary ileostomy. Pathology indicated moderately active chronic ulcerative colitis, involving colonic mucosa from the distal margin to the ascending colon. Family history included hyperlipidemia, depressive disorder, malignant cervical neoplasm, viral hepatitis A, and depressive disorder.
TESTNOT17	pINCY	Library was constructed from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor).
THP1TXT03	pINCY	Library was constructed using RNA isolated from treated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year- old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171). The THP-1 cultured cells were differentiated with PMA(100ng/ml) for 48 hours, incubated with Mycobacteria tuberculosis, strain H37Rv, for 4 hours at 37C, washed and RNA extracted.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
50	7511620	1733477T6	SNP00099024	413	2028	A	A	G	noncoding	n/a	n/a	n/a	n/a
50	7511620	2416475T6	SNP00099024	430	2039	A	A	G	noncoding	n/a	n/a	n/a	n/a
50	7511620	7685191H1	SNP00099024	348	2027	G	A	G	noncoding	n/a	n/a	n/a	n/a
51	7506995	1878368H1	SNP00034867	143	2120	T	C	T	noncoding	1.00	n/a	n/a	n/a
51	7506995	2601563F6	SNP00039549	158	1463	G	G	A	V390	0.17	n/a	n/a	n/a
51	7506995	2753227H1	SNP00010061	132	1871	A	A	G	noncoding	n/a	n/a	n/a	n/a
51	7506995	3051974H1	SNP00039548	25	1181	T	G	T	A296	n/a	n/a	n/a	n/a
51	7506995	5884565H1	SNP00147613	53	1181	G	G	C	A296	n/a	n/a	n/a	n/a
52	7506996	1293671H1	SNP00039547	174	819	G	G	A	G175	n/a	n/a	n/a	n/a
52	7506996	1878368H1	SNP00034867	143	2191	T	C	T	noncoding	1.00	n/a	n/a	n/a
52	7506996	2601563F6	SNP00039549	158	1534	G	G	A	V413	0.17	n/a	n/a	n/a
52	7506996	2753227H1	SNP00010061	132	1942	A	A	G	noncoding	n/a	n/a	n/a	n/a
52	7506996	3051974H1	SNP00039548	25	1252	T	G	T	A319	n/a	n/a	n/a	n/a
52	7506996	5884565H1	SNP00147613	53	1252	G	G	C	A319	n/a	n/a	n/a	n/a
52	7506996	6060785H1	SNP00135254	234	786	A	G	A	K164	n/a	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, 11, 14-18, 20, 22-23, and 25-26,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 93% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:9 and SEQ ID NO:21,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:24,

e) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:19,

f) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and

g) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-44, SEQ ID NO:46-47, SEQ ID NO:49, and SEQ ID NO:51-52,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 92% identical to the polynucleotide sequence of SEQ ID NO:48,

d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to the polynucleotide sequence of SEQ ID NO:45,

e) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:50,

d) a polynucleotide complementary to a polynucleotide of a),

e) a polynucleotide complementary to a polynucleotide of b),

f) a polynucleotide complementary to a polynucleotide of c),

g) a polynucleotide complementary to a polynucleotide of d),

h) a polynucleotide complementary to a polynucleotide of e), and

i) an RNA equivalent of a)-h).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

19. A method for treating a disease or condition associated with decreased expression of

functional TRICH, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is

indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of TRICH in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, or an immunogenic fragment

thereof, under conditions to elicit an antibody response,

b) isolating antibodies from the animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is

completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:51.

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:52.

<110> INCYTE GENOMICS, INC.
TANG, Y. Tom
LAL, Preeti G.
YUE, Henry
BAUGHN, Mariah R.
NGUYEN, Danniel B.
YAO, Monique G.
GREENE, Barrie D.
BOROWSKY, Mark L.
LEE, Sally
EMERLING, Brooke M.
XU, Yuming
BECHA, Shanya D.
GORVAD, Ann E.
AZIMZAI, Yalda
YUE, Huibin
ELLIOTT, Vicki S.
LEE, Ernestine A.
YANG, Junming
LEHR-MASON, Patricia M.
RAMKUMAR, Jayalaxmi
LEE, Soo Yeun
FARIS, Mary
TURNER, Christopher
FURNESS, Michael
BUCHBINDER, Jenny L.
WALIA, Narinder K.
LI, Joana X
FORSYTHE, Ian J.
GRIFFIN, Jennifer A.
GIETZEN, Kimberly J.
SWARNAKAR, Anita
HAFALIA, April J.A.
LINDQUIST, Erika A.
JIANG, Xin
JACKSON, Alan A.
WILSON, Amy D.
JIN, Pei
KHARE, Reena
MARQUIS, Joseph P.

<120> TRANSPORTERS AND ION CHANNELS

<130> PF-1148 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/313,242
<151> 2001-08-17

<150> 60/324,782
<151> 2001-09-21

<150> 60/328,184
<151> 2001-10-02

<150> 60/345,937
<151> 2001-10-26

<150> 60/335,698
<151> 2001-11-01

<150> 60/332,804
<151> 2001-11-13

<150> 60/333,922
<151> 2001-11-27

<150> 60/375,637
<151> 2002-04-26

<150> 60/377,444
<151> 2002-05-03

<150> 60/388,180
<151> 2002-06-11

<160> 52
<170> PERL Program

<210> 1
<211> 652
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1853191CD1

<400> 1

```
Met His Cys Leu Gly Ala Glu Tyr Leu Val Ser Ala Glu Gly Ala
  1               5                  10                  15
Pro Arg Gln Arg Glu Trp Arg Pro Gln Ile Tyr Arg Lys Cys Thr
             20                  25                  30
Asp Thr Ala Trp Leu Phe Leu Phe Phe Leu Phe Trp Thr Gly Leu
             35                  40                  45
Val Phe Ile Met Gly Tyr Ser Val Val Ala Gly Ala Ala Gly Arg
             50                  55                  60
Leu Leu Phe Gly Tyr Asp Ser Phe Gly Asn Met Cys Gly Lys Lys
             65                  70                  75
Asn Ser Pro Val Glu Gly Ala Pro Leu Ser Gly Gln Asp Met Thr
             80                  85                  90
Leu Lys Lys His Val Phe Phe Met Asn Ser Cys Asn Leu Glu Val
             95                 100                 105
Lys Gly Thr Gln Leu Asn Arg Met Ala Leu Cys Val Ser Asn Cys
            110                 115                 120
Pro Glu Glu Gln Leu Asp Ser Leu Glu Glu Val Gln Phe Phe Ala
            125                 130                 135
Asn Thr Ser Gly Ser Phe Leu Cys Val Tyr Ser Leu Asn Ser Phe
            140                 145                 150
Asn Tyr Thr His Ser Pro Lys Ala Asp Ser Leu Cys Pro Arg Leu
            155                 160                 165
```

```
Pro Val Pro Pro Ser Lys Ser Phe Pro Leu Phe Asn Arg Cys Val
                170                 175                 180
Pro Gln Thr Pro Glu Cys Tyr Ser Leu Phe Ala Ser Val Leu Ile
                185                 190                 195
Asn Asp Val Asp Thr Leu His Arg Ile Leu Ser Gly Ile Met Ser
                200                 205                 210
Gly Arg Asp Thr Ile Leu Gly Leu Cys Ile Leu Ala Leu Ala Leu
                215                 220                 225
Ser Leu Ala Met Met Phe Thr Phe Arg Phe Ile Thr Thr Leu Leu
                230                 235                 240
Val His Ile Phe Ile Ser Leu Val Ile Leu Gly Leu Leu Phe Val
                245                 250                 255
Cys Gly Val Leu Trp Trp Leu Tyr Tyr Asp Tyr Thr Asn Asp Leu
                260                 265                 270
Ser Ile Glu Leu Asp Thr Glu Arg Glu Asn Met Lys Cys Val Leu
                275                 280                 285
Gly Phe Ala Ile Val Ser Thr Gly Ile Thr Ala Val Leu Leu Val
                290                 295                 300
Leu Ile Phe Val Leu Arg Lys Arg Ile Lys Leu Thr Val Glu Leu
                305                 310                 315
Phe Gln Ile Thr Asn Lys Ala Ile Ser Ser Ala Pro Phe Leu Leu
                320                 325                 330
Phe Gln Pro Leu Trp Thr Phe Ala Ile Leu Ile Phe Phe Trp Val
                335                 340                 345
Leu Trp Val Ala Val Leu Leu Ser Leu Gly Thr Ala Gly Ala Ala
                350                 355                 360
Gln Val Met Glu Gly Gly Gln Val Glu Tyr Lys Pro Leu Ser Gly
                365                 370                 375
Ile Arg Tyr Met Trp Ser Tyr His Leu Ile Gly Leu Ile Trp Thr
                380                 385                 390
Ser Glu Phe Ile Leu Ala Cys Gln Gln Met Thr Ile Ala Gly Ala
                395                 400                 405
Val Val Thr Cys Tyr Phe Asn Arg Ser Lys Asn Asp Pro Pro Asp
                410                 415                 420
His Pro Ile Leu Ser Ser Leu Ser Ile Leu Phe Phe Tyr His Gln
                425                 430                 435
Gly Thr Ile Val Lys Gly Ser Phe Leu Ile Ser Val Val Arg Ile
                440                 445                 450
Pro Arg Ile Ile Val Met Tyr Met Gln Asn Ala Leu Lys Glu Gln
                455                 460                 465
His Gly Ala Leu Ser Arg Tyr Leu Phe Arg Cys Cys Tyr Cys Cys
                470                 475                 480
Phe Trp Cys Leu Asp Lys Tyr Leu Leu His Leu Asn Gln Asn Ala
                485                 490                 495
Tyr Thr Thr Thr Ala Ile Asn Gly Thr Asp Phe Cys Thr Ser Ala
                500                 505                 510
Lys Asp Ala Phe Lys Ile Leu Ser Lys Asn Ser Ser His Phe Thr
                515                 520                 525
Ser Ile Asn Cys Phe Gly Asp Phe Ile Ile Phe Leu Gly Lys Val
                530                 535                 540
Leu Val Val Cys Phe Thr Val Phe Gly Gly Leu Met Ala Phe Asn
                545                 550                 555
Tyr Asn Arg Ala Phe Gln Val Trp Ala Val Pro Leu Leu Leu Val
                560                 565                 570
Ala Phe Phe Ala Tyr Leu Val Ala His Ser Phe Leu Ser Val Phe
                575                 580                 585
```

```
Glu Thr Val Leu Asp Ala Leu Phe Leu Cys Phe Ala Val Asp Leu
                590                 595                 600
Glu Thr Asn Asp Gly Ser Ser Glu Lys Pro Tyr Phe Met Asp Gln
                605                 610                 615
Glu Phe Leu Ser Phe Val Lys Arg Ser Asn Lys Leu Asn Asn Ala
                620                 625                 630
Arg Ala Gln Gln Asp Lys His Ser Leu Arg Asn Glu Glu Gly Thr
                635                 640                 645
Glu Leu Gln Ala Ile Val Arg
                650
```

<210> 2
<211> 345
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7497369CD1

<400> 2

```
Met Pro Gly Lys Arg Met Gly Tyr Arg Val Thr Glu Asn Tyr Ser
  1                 5                  10                  15
Asp Val Leu Pro Leu Ser Leu Gly Val Gly Trp Gly Pro Trp Pro
                 20                  25                  30
Leu Pro Leu Leu Pro Gly Gly Gly Gly Gly Ser Arg Gly His Pro
                 35                  40                  45
His Gln Arg Cys Arg Phe Pro Ala Leu Phe Pro Glu Ser Pro Cys
                 50                  55                  60
Trp Leu Leu Ala Thr Gly Gln Val Ala Arg Ala Arg Lys Ile Leu
                 65                  70                  75
Trp Arg Phe Ala Glu Ala Ser Gly Val Gly Pro Gly Asp Ser Pro
                 80                  85                  90
Leu Glu Glu Asn Ser Leu Ala Thr Glu Leu Thr Met Leu Ser Ala
                 95                 100                 105
Arg Ser Pro Gln Pro Arg Tyr His Ser Pro Leu Gly Leu Leu Arg
                110                 115                 120
Thr Arg Val Thr Trp Arg Asn Gly Leu Ile Leu Gly Phe Ser Ser
                125                 130                 135
Leu Val Gly Gly Gly Ile Arg Ala Ser Phe Arg Arg Ser Leu Ala
                140                 145                 150
Pro Gln Val Pro Thr Phe Tyr Leu Pro Tyr Phe Leu Glu Ala Gly
                155                 160                 165
Leu Glu Ala Ala Ala Leu Val Phe Leu Leu Leu Thr Ala Asp Cys
                170                 175                 180
Cys Gly Arg Arg Pro Val Leu Leu Leu Gly Thr Met Val Thr Gly
                185                 190                 195
Leu Ala Ser Leu Leu Leu Leu Ala Gly Ala Gln Tyr Leu Pro Gly
                200                 205                 210
Trp Thr Val Leu Phe Leu Ser Val Leu Gly Leu Leu Ala Ser Arg
                215                 220                 225
Ala Val Ser Ala Leu Ser Ser Leu Phe Ala Ala Glu Val Phe Pro
                230                 235                 240
Thr Val Ile Arg Gly Ala Gly Leu Gly Leu Val Leu Gly Ala Gly
                245                 250                 255
Phe Leu Gly Gln Ala Ala Gly Pro Leu Asp Thr Leu His Gly Arg
```

```
                260                 265                 270
Gln Gly Phe Phe Leu Gln Gln Val Val Phe Ala Ser Leu Ala Val
                275                 280                 285
Leu Ala Leu Leu Cys Val Leu Leu Leu Pro Glu Ser Arg Ser Arg
                290                 295                 300
Gly Leu Pro Gln Ser Leu Gln Asp Ala Asp Arg Leu Arg Arg Ser
                305                 310                 315
Pro Leu Leu Arg Gly Arg Pro Arg Gln Asp His Leu Pro Leu Leu
                320                 325                 330
Pro Pro Ser Asn Ser Tyr Trp Ala Gly His Thr Pro Glu Gln His
                335                 340                 345


<210> 3
<211> 150
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1700438CD1

<400> 3
Met Gln Arg Leu Leu Cys Met Cys His Glu Gly Trp Glu Ala Tyr
  1               5                  10                  15
Cys Arg Gln Met Val Phe Leu Ala Gly Leu Cys Leu Val Phe Leu
                20                  25                  30
Tyr Met Thr Val Leu Gly Ser Gly Gly Ile Ile Thr Gly Tyr Ala
                35                  40                  45
Cys Thr Gln Gly Val Gly Asp Ser Leu Leu Ser Ile Leu Thr Ala
                50                  55                  60
Leu Ser Ala Leu Ser Gly Leu Met Gly Thr Val Leu Phe Thr Gln
                65                  70                  75
Leu Arg Gly His Tyr Gly Leu Val Thr Thr Gly Val Ile Ser Ser
                80                  85                  90
Gln Leu His Leu Gly Cys Leu Met Leu Cys Met Phe Ser Val Leu
                95                 100                 105
Ala Pro Gly Asn Ser Phe Asp Leu Ala Val Phe Ser Leu Pro Leu
               110                 115                 120
Ser Lys Asn Pro Ser Asn Tyr Glu Leu Leu Val Gln Trp Met Glu
               125                 130                 135
Glu Gln Ser Arg Gly Met Ala Trp Phe Arg Phe Leu Ser Lys Gly
               140                 145                 150


<210> 4
<211> 537
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 535939CD1

<400> 4
Met Gly Asp Glu Asp Lys Arg Ile Thr Tyr Glu Asp Ser Glu Pro
```

```
  1               5                  10                  15
Ser Thr Gly Met Asn Tyr Thr Pro Ser Met His Gln Glu Ala Gln
                20                  25                  30
Glu Glu Thr Val Met Lys Leu Lys Gly Ile Asp Ala Asn Glu Pro
                35                  40                  45
Thr Glu Gly Ser Ile Leu Leu Lys Ser Ser Glu Lys Lys Leu Gln
                50                  55                  60
Glu Thr Pro Thr Glu Ala Asn His Val Gln Arg Leu Arg Gln Met
                65                  70                  75
Leu Ala Cys Pro Pro His Gly Leu Leu Asp Arg Val Ile Thr Asn
                80                  85                  90
Val Thr Ile Ile Val Leu Leu Trp Ala Val Val Trp Ser Ile Thr
                95                 100                 105
Gly Ser Glu Cys Leu Pro Gly Gly Asn Leu Phe Gly Ile Ile Ile
               110                 115                 120
Leu Phe Tyr Cys Ala Ile Ile Gly Gly Lys Leu Leu Gly Leu Ile
               125                 130                 135
Lys Leu Pro Thr Leu Pro Pro Leu Pro Ser Leu Leu Gly Met Leu
               140                 145                 150
Leu Ala Gly Phe Leu Ile Arg Asn Ile Pro Val Ile Asn Asp Asn
               155                 160                 165
Val Gln Ile Lys His Lys Trp Ser Ser Ser Leu Arg Ser Ile Ala
               170                 175                 180
Leu Ser Ile Ile Leu Val Arg Ala Gly Leu Gly Leu Asp Ser Lys
               185                 190                 195
Ala Leu Lys Lys Leu Lys Gly Val Cys Val Arg Leu Ser Met Gly
               200                 205                 210
Pro Cys Ile Val Glu Ala Cys Thr Ser Ala Leu Leu Ala His Tyr
               215                 220                 225
Leu Leu Gly Leu Pro Trp Gln Trp Gly Phe Ile Leu Gly Phe Val
               230                 235                 240
Leu Gly Ala Val Ser Pro Ala Val Val Val Pro Ser Met Leu Leu
               245                 250                 255
Leu Gln Gly Gly Gly Tyr Gly Val Glu Lys Gly Val Pro Thr Leu
               260                 265                 270
Leu Met Ala Ala Gly Ser Phe Asp Asp Ile Leu Ala Ile Thr Gly
               275                 280                 285
Phe Asn Thr Cys Leu Gly Ile Ala Phe Ser Thr Gly Ser Thr Val
               290                 295                 300
Phe Asn Val Leu Arg Gly Val Leu Glu Val Val Ile Gly Val Ala
               305                 310                 315
Thr Gly Ser Val Leu Gly Phe Phe Ile Gln Tyr Phe Pro Ser Arg
               320                 325                 330
Asp Gln Asp Lys Leu Val Cys Lys Arg Thr Phe Leu Val Leu Gly
               335                 340                 345
Leu Ser Val Leu Ala Val Phe Ser Ser Val His Phe Gly Phe Pro
               350                 355                 360
Gly Ser Gly Gly Leu Cys Thr Leu Val Met Ala Phe Leu Ala Gly
               365                 370                 375
Met Gly Trp Thr Ser Glu Lys Ala Glu Val Glu Lys Ile Ile Ala
               380                 385                 390
Val Ala Trp Asp Ile Phe Gln Pro Leu Leu Phe Gly Leu Ile Gly
               395                 400                 405
Ala Glu Val Ser Ile Ala Ser Leu Arg Pro Glu Thr Val Gly Leu
               410                 415                 420
Cys Val Ala Thr Val Gly Ile Ala Val Leu Ile Arg Ile Leu Thr
```

```
                425                 430                 435
Thr Phe Leu Met Val Cys Phe Ala Gly Phe Asn Leu Lys Glu Lys
                440                 445                 450
Ile Phe Ile Ser Phe Ala Trp Leu Pro Lys Ala Thr Val Gln Ala
                455                 460                 465
Ala Ile Gly Ser Val Ala Leu Asp Thr Ala Arg Ser His Gly Glu
                470                 475                 480
Lys Gln Leu Glu Asp Tyr Gly Met Asp Val Leu Thr Val Ala Phe
                485                 490                 495
Leu Ser Ile Leu Ile Thr Ala Pro Ile Gly Ser Leu Leu Ile Gly
                500                 505                 510
Leu Leu Gly Pro Arg Leu Leu Gln Lys Val Glu His Gln Asn Lys
                515                 520                 525
Asp Glu Glu Val Gln Gly Glu Thr Ser Val Gln Val
                530                 535

<210> 5
<211> 1119
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55118067CD1

<400> 5
Met Thr Ala Ala Ala Ala Ser Asn Trp Gly Leu Ile Thr Asn Ile
  1               5                  10                  15
Val Asn Ser Ile Val Gly Val Ser Val Leu Thr Met Pro Phe Cys
                20                  25                  30
Phe Lys Gln Cys Gly Ile Val Leu Gly Ala Leu Leu Leu Val Phe
                35                  40                  45
Cys Ser Trp Met Thr His Gln Ser Cys Met Phe Leu Val Lys Ser
                50                  55                  60
Ala Ser Leu Ser Lys Arg Arg Thr Tyr Ala Gly Leu Ala Phe His
                65                  70                  75
Ala Tyr Gly Lys Ala Gly Lys Met Leu Val Glu Thr Ser Met Ile
                80                  85                  90
Gly Leu Met Leu Gly Thr Cys Ile Ala Phe Tyr Val Val Ile Gly
                95                 100                 105
Asp Leu Gly Ser Asn Phe Phe Ala Arg Leu Phe Gly Phe Gln Val
                110                 115                 120
Gly Gly Thr Phe Arg Met Phe Leu Leu Phe Ala Val Ser Leu Cys
                125                 130                 135
Ile Val Leu Pro Leu Ser Leu Gln Arg Asn Met Met Ala Ser Ile
                140                 145                 150
Gln Ser Phe Ser Ala Met Ala Leu Leu Phe Tyr Thr Val Phe Met
                155                 160                 165
Phe Val Ile Val Leu Ser Ser Leu Lys His Gly Leu Phe Ser Gly
                170                 175                 180
Gln Trp Leu Arg Arg Val Ser Tyr Val Arg Trp Glu Gly Val Phe
                185                 190                 195
Arg Cys Ile Pro Ile Phe Gly Met Ser Phe Ala Cys Gln Ser Gln
                200                 205                 210
Val Leu Pro Thr Tyr Asp Ser Leu Asp Glu Pro Ser Val Lys Thr
                215                 220                 225
```

```
Met Ser Ser Ile Phe Ala Ser Ser Leu Asn Val Val Thr Thr Phe
                230                 235                 240
Tyr Val Met Val Gly Phe Phe Gly Tyr Val Ser Phe Thr Glu Ala
                245                 250                 255
Thr Ala Gly Asn Val Leu Met His Phe Pro Ser Asn Leu Val Thr
                260                 265                 270
Glu Met Leu Arg Val Gly Phe Met Met Ser Val Ala Val Gly Phe
                275                 280                 285
Pro Met Met Ile Leu Pro Cys Arg Gln Ala Leu Ser Thr Leu Leu
                290                 295                 300
Cys Glu Gln Gln Gln Lys Asp Gly Thr Phe Ala Ala Gly Gly Tyr
                305                 310                 315
Met Pro Pro Leu Arg Phe Lys Ala Leu Thr Leu Ser Val Val Phe
                320                 325                 330
Gly Thr Met Val Gly Gly Ile Leu Ile Pro Asn Val Glu Thr Ile
                335                 340                 345
Leu Gly Leu Thr Gly Ala Thr Met Gly Ser Leu Ile Cys Phe Ile
                350                 355                 360
Cys Pro Ala Leu Ile Tyr Lys Lys Ile His Lys Asn Ala Leu Ser
                365                 370                 375
Ser Gln Val Val Leu Trp Val Gly Leu Gly Val Leu Val Val Ser
                380                 385                 390
Thr Val Thr Thr Leu Ser Val Ser Glu Glu Val Pro Glu Asp Leu
                395                 400                 405
Ala Glu Glu Ala Pro Gly Gly Arg Leu Gly Glu Ala Glu Gly Leu
                410                 415                 420
Met Lys Val Glu Ala Ala Arg Leu Ser Ala Gln Asp Pro Val Val
                425                 430                 435
Ala Val Ala Glu Asp Gly Arg Glu Lys Pro Lys Leu Pro Lys Glu
                440                 445                 450
Arg Glu Glu Leu Glu Gln Ala Gln Ile Lys Gly Pro Val Asp Val
                455                 460                 465
Pro Gly Arg Glu Asp Gly Lys Glu Ala Pro Glu Glu Ala Gln Leu
                470                 475                 480
Asp Arg Pro Gly Gln Gly Ile Ala Val Pro Val Gly Glu Ala His
                485                 490                 495
Arg His Glu Pro Pro Val Pro His Asp Lys Val Val Val Asp Glu
                500                 505                 510
Gly Gln Asp Arg Glu Val Pro Glu Glu Asn Lys Pro Pro Ser Arg
                515                 520                 525
His Ala Gly Gly Lys Ala Pro Gly Val Gln Gly Gln Met Ala Pro
                530                 535                 540
Pro Leu Pro Asp Ser Glu Arg Glu Lys Gln Glu Pro Glu Gln Gly
                545                 550                 555
Glu Val Gly Lys Arg Pro Gly Gln Ala Gln Ala Leu Glu Glu Ala
                560                 565                 570
Gly Asp Leu Pro Glu Asp Pro Gln Lys Val Pro Glu Ala Asp Gly
                575                 580                 585
Gln Pro Ala Val Gln Pro Ala Lys Glu Asp Leu Gly Pro Gly Asp
                590                 595                 600
Arg Gly Leu His Pro Arg Pro Gln Ala Val Leu Ser Glu Gln Gln
                605                 610                 615
Asn Gly Leu Ala Val Gly Gly Gly Glu Lys Ala Lys Gly Gly Pro
                620                 625                 630
Pro Pro Gly Asn Ala Ala Gly Asp Thr Gly Gln Pro Ala Glu Asp
                635                 640                 645
```

```
Ser Asp His Gly Gly Lys Pro Pro Leu Pro Ala Glu Lys Pro Ala
                650                 655                 660
Pro Gly Pro Gly Leu Pro Pro Glu Pro Arg Glu Gln Arg Asp Val
                665                 670                 675
Glu Arg Ala Gly Gly Asn Gln Ala Ala Ser Gln Leu Glu Glu Ala
                680                 685                 690
Gly Arg Ala Glu Met Leu Asp His Ala Val Leu Leu Gln Val Ile
                695                 700                 705
Lys Glu Gln Gln Val Gln Gln Lys Arg Leu Leu Asp Gln Gln Glu
                710                 715                 720
Lys Leu Leu Ala Val Ile Glu Glu Gln His Lys Glu Ile His Gln
                725                 730                 735
Gln Arg Gln Glu Asp Glu Glu Asp Lys Pro Arg Gln Val Glu Val
                740                 745                 750
His Gln Glu Pro Gly Ala Ala Val Pro Arg Gly Gln Glu Ala Pro
                755                 760                 765
Glu Gly Lys Ala Arg Glu Thr Val Glu Asn Leu Pro Pro Leu Pro
                770                 775                 780
Leu Asp Pro Val Leu Arg Ala Pro Gly Gly Arg Pro Ala Pro Ser
                785                 790                 795
Gln Asp Leu Asn Gln Arg Ser Leu Glu His Ser Glu Gly Pro Val
                800                 805                 810
Gly Arg Asp Pro Ala Gly Pro Pro Asp Gly Gly Pro Asp Thr Glu
                815                 820                 825
Pro Arg Ala Ala Gln Ala Lys Leu Arg Asp Gly Gln Lys Asp Ala
                830                 835                 840
Ala Pro Arg Ala Ala Gly Thr Val Lys Glu Leu Pro Lys Gly Pro
                845                 850                 855
Glu Gln Val Pro Val Pro Asp Pro Ala Arg Glu Ala Gly Gly Pro
                860                 865                 870
Glu Glu Arg Leu Ala Glu Glu Phe Pro Gly Gln Ser Gln Asp Val
                875                 880                 885
Thr Gly Gly Ser Gln Asp Arg Lys Lys Pro Gly Lys Glu Val Ala
                890                 895                 900
Ala Thr Gly Thr Ser Ile Leu Lys Glu Ala Asn Trp Leu Val Ala
                905                 910                 915
Gly Pro Gly Ala Glu Thr Gly Asp Pro Arg Met Lys Pro Lys Gln
                920                 925                 930
Val Ser Arg Asp Leu Gly Leu Ala Ala Asp Leu Pro Gly Gly Ala
                935                 940                 945
Glu Gly Ala Ala Ala Gln Pro Gln Ala Val Leu Arg Gln Pro Glu
                950                 955                 960
Leu Arg Val Ile Ser Asp Gly Glu Gln Gly Gly Gln Gln Gly His
                965                 970                 975
Arg Leu Asp His Gly Gly His Leu Glu Met Arg Lys Ala Arg Gly
                980                 985                 990
Gly Asp His Val Pro Val Ser His Glu Gln Pro Arg Gly Gly Glu
                995                1000                1005
Asp Ala Ala Val Gln Glu Pro Arg Gln Arg Pro Glu Pro Glu Leu
               1010                1015                1020
Gly Leu Lys Arg Ala Val Pro Gly Gly Gln Arg Pro Asp Asn Ala
               1025                1030                1035
Lys Pro Asn Arg Asp Leu Lys Leu Gln Ala Gly Ser Asp Leu Arg
               1040                1045                1050
Arg Arg Arg Arg Asp Leu Gly Pro His Ala Glu Gly Gln Leu Ala
               1055                1060                1065
```

```
Pro Arg Asp Gly Val Ile Ile Gly Leu Asn Pro Leu Pro Asp Val
                1070                1075                1080
Gln Val Asn Asp Leu Arg Gly Ala Leu Asp Ala Gln Leu Arg Gln
                1085                1090                1095
Ala Ala Gly Gly Ala Leu Gln Val Val His Ser Arg Gln Leu Arg
                1100                1105                1110
Gln Ala Pro Gly Pro Pro Glu Glu Ser
                1115
```

<210> 6
<211> 947
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7502087CD1

<400> 6

```
Met Pro Val Arg Arg Gly His Val Ala Pro Gln Asn Thr Tyr Leu
  1               5                  10                  15
Asp Thr Ile Ile Arg Lys Phe Glu Gly Gln Ser Arg Lys Phe Leu
                20                  25                  30
Ile Ala Asn Ala Gln Met Glu Asn Cys Ala Ile Ile Tyr Cys Asn
                35                  40                  45
Asp Gly Phe Cys Glu Leu Phe Gly Tyr Ser Arg Val Glu Val Met
                50                  55                  60
Gln Gln Pro Cys Thr Cys Asp Phe Leu Thr Gly Pro Asn Thr Pro
                65                  70                  75
Ser Ser Ala Val Ser Arg Leu Ala Gln Ala Leu Leu Gly Ala Glu
                80                  85                  90
Glu Cys Lys Val Asp Ile Leu Tyr Tyr Arg Lys Asp Ala Ser Ser
                95                 100                 105
Phe Arg Cys Leu Val Asp Val Val Pro Val Lys Asn Glu Asp Gly
               110                 115                 120
Ala Val Ile Met Phe Ile Leu Asn Phe Glu Asp Leu Ala Gln Leu
               125                 130                 135
Leu Ala Lys Cys Ser Ser Arg Ser Leu Ser Gln Arg Leu Leu Ser
               140                 145                 150
Gln Ser Phe Leu Gly Ser Glu Gly Ser His Gly Arg Pro Gly Gly
               155                 160                 165
Pro Gly Pro Gly Thr Gly Arg Gly Lys Tyr Arg Thr Ile Ser Gln
               170                 175                 180
Ile Pro Gln Phe Thr Leu Asn Phe Val Glu Phe Asn Leu Glu Lys
               185                 190                 195
His Arg Ser Ser Ser Thr Thr Glu Ile Glu Ile Ile Ala Pro His
               200                 205                 210
Lys Val Val Glu Arg Thr Gln Asn Val Thr Glu Lys Val Thr Gln
               215                 220                 225
Val Leu Ser Leu Gly Ala Asp Val Leu Pro Glu Tyr Lys Leu Gln
               230                 235                 240
Ala Pro Arg Ile His Arg Trp Thr Ile Leu His Tyr Ser Pro Phe
               245                 250                 255
Lys Ala Val Trp Asp Trp Leu Ile Leu Leu Leu Val Ile Tyr Thr
               260                 265                 270
Ala Val Phe Thr Pro Tyr Ser Ala Ala Phe Leu Leu Ser Asp Gln
```

```
                275                 280                 285
Asp Glu Ser Arg Arg Gly Ala Cys Ser Tyr Thr Cys Ser Pro Leu
                290                 295                 300
Thr Val Val Asp Leu Ile Val Asp Ile Met Phe Val Val Asp Ile
                305                 310                 315
Val Ile Asn Phe Arg Thr Thr Tyr Val Asn Thr Asn Asp Glu Val
                320                 325                 330
Val Ser His Pro Arg Arg Ile Ala Val His Tyr Phe Lys Gly Trp
                335                 340                 345
Phe Leu Ile Asp Met Val Ala Ala Ile Pro Phe Asp Leu Leu Ile
                350                 355                 360
Phe Arg Thr Gly Ser Asp Glu Thr Thr Thr Leu Ile Gly Leu Leu
                365                 370                 375
Lys Thr Ala Arg Leu Leu Arg Leu Val Arg Val Ala Arg Lys Leu
                380                 385                 390
Asp Cys Tyr Ser Glu Tyr Gly Ala Ala Val Leu Phe Leu Leu Met
                395                 400                 405
Cys Thr Phe Ala Leu Ile Ala His Trp Leu Ala Cys Ile Trp Tyr
                410                 415                 420
Ala Ile Gly Asn Val Glu Arg Pro Tyr Leu Glu His Lys Ile Gly
                425                 430                 435
Trp Leu Asp Ser Leu Gly Val Gln Leu Gly Lys Arg Tyr Asn Gly
                440                 445                 450
Ser Asp Pro Ala Ser Gly Pro Ser Val Gln Asp Lys Tyr Val Thr
                455                 460                 465
Ala Leu Tyr Phe Thr Phe Ser Ser Leu Thr Ser Val Gly Phe Gly
                470                 475                 480
Asn Val Ser Pro Asn Thr Asn Ser Glu Lys Val Phe Ser Ile Cys
                485                 490                 495
Val Met Leu Ile Gly Ser Leu Met Tyr Ala Ser Ile Phe Gly Asn
                500                 505                 510
Val Ser Ala Ile Ile Gln Arg Leu Tyr Ser Gly Thr Ala Arg Tyr
                515                 520                 525
His Thr Gln Met Leu Arg Val Lys Glu Phe Ile Arg Phe His Gln
                530                 535                 540
Ile Pro Asn Pro Leu Arg Gln Arg Leu Glu Glu Tyr Phe Gln His
                545                 550                 555
Ala Trp Ser Tyr Thr Asn Gly Ile Asp Met Asn Ala Val Leu Lys
                560                 565                 570
Gly Phe Pro Glu Cys Leu Gln Ala Asp Ile Cys Leu His Leu His
                575                 580                 585
Arg Ala Leu Leu Gln His Cys Pro Ala Phe Ser Gly Ala Gly Lys
                590                 595                 600
Gly Cys Leu Arg Ala Leu Ala Val Lys Phe Lys Thr Thr His Ala
                605                 610                 615
Pro Pro Gly Asp Thr Leu Val His Leu Gly Asp Val Leu Ser Thr
                620                 625                 630
Leu Tyr Phe Ile Ser Arg Gly Ser Ile Glu Ile Leu Arg Asp Asp
                635                 640                 645
Val Val Val Ala Ile Leu Gly Lys Asn Asp Ile Phe Gly Glu Pro
                650                 655                 660
Val Ser Leu His Ala Gln Pro Gly Lys Ser Ser Ala Asp Val Arg
                665                 670                 675
Ala Leu Thr Tyr Cys Asp Leu His Lys Ile Gln Arg Ala Asp Leu
                680                 685                 690
Leu Glu Val Leu Asp Met Tyr Pro Ala Phe Ala Glu Ser Phe Trp
```

695 700 705
Ser Lys Leu Glu Val Thr Phe Asn Leu Arg Asp Ala Pro Gly Ser
710 715 720
Gln Asp His Gln Gly Phe Phe Leu Ser Asp Asn Gln Ser Asp Ala
725 730 735
Ala Pro Pro Leu Ser Ile Ser Asp Ala Ser Gly Leu Trp Pro Glu
740 745 750
Leu Leu Gln Glu Met Pro Pro Arg His Ser Pro Gln Ser Pro Gln
755 760 765
Glu Asp Pro Asp Cys Trp Pro Leu Lys Leu Gly Ser Arg Leu Glu
770 775 780
Gln Leu Gln Ala Gln Met Asn Arg Leu Glu Ser Arg Val Ser Ser
785 790 795
Asp Leu Ser Arg Ile Leu Gln Leu Leu Gln Lys Pro Met Pro Gln
800 805 810
Gly His Ala Ser Tyr Ile Leu Glu Ala Pro Ala Ser Asn Asp Leu
815 820 825
Ala Leu Val Pro Ile Ala Ser Glu Thr Thr Ser Pro Gly Pro Arg
830 835 840
Leu Pro Gln Gly Phe Leu Pro Pro Ala Gln Thr Pro Ser Tyr Gly
845 850 855
Asp Leu Asp Asp Cys Ser Pro Lys His Arg Asn Ser Ser Pro Arg
860 865 870
Met Pro His Leu Ala Val Ala Met Asp Lys Thr Leu Ala Pro Ser
875 880 885
Ser Glu Gln Glu Gln Pro Glu Gly Leu Trp Pro Pro Leu Ala Ser
890 895 900
Pro Leu His Pro Leu Glu Val Gln Gly Leu Ile Cys Gly Pro Cys
905 910 915
Phe Ser Ser Leu Pro Glu His Leu Gly Ser Val Pro Lys Gln Leu
920 925 930
Asp Phe Gln Arg His Gly Ser Asp Pro Gly Phe Ala Gly Ser Trp
935 940 945
Gly His

<210> 7
<211> 80
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500819CD1

<400> 7
Met Glu Leu Val Leu Val Phe Leu Cys Ser Leu Leu Ala Pro Met
1 5 10 15
Val Leu Ala Ser Ala Ala Glu Lys Glu Lys Glu Met Asp Pro Phe
20 25 30
His Tyr Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu Val Phe Ala
35 40 45
Val Val Leu Phe Ser Val Gly Ile Leu Leu Ile Leu Ser Arg Arg
50 55 60
Cys Lys Cys Ser Phe Tyr Ser Ala Pro Gly Glu Cys Val Pro Cys
65 70 75

Ile Ser Ser Gln Gln
80

<210> 8
<211> 531
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503413CD1

<400> 8
Met Ala Ser Ala Leu Ser Tyr Val Ser Lys Phe Lys Ser Phe Val
1 5 10 15
Ile Leu Phe Val Thr Pro Leu Leu Leu Leu Pro Leu Val Ile Leu
20 25 30
Met Pro Ala Lys Phe Val Arg Cys Ala Tyr Val Ile Ile Leu Met
35 40 45
Ala Ile Tyr Trp Cys Thr Glu Val Ile Pro Leu Ala Val Thr Ser
50 55 60
Leu Met Pro Val Leu Leu Phe Pro Leu Phe Gln Ile Leu Asp Ser
65 70 75
Arg Gln Val Cys Val Gln Tyr Met Lys Asp Thr Asn Met Leu Phe
80 85 90
Leu Gly Gly Leu Ile Val Ala Val Ala Val Glu Arg Trp Asn Leu
95 100 105
His Lys Arg Ile Ala Leu Arg Thr Leu Leu Trp Val Gly Ala Lys
110 115 120
Pro Ala Arg Leu Met Leu Gly Phe Met Gly Val Thr Ala Leu Leu
125 130 135
Ser Met Trp Ile Ser Asn Thr Ala Thr Thr Ala Met Met Val Pro
140 145 150
Ile Val Glu Ala Ile Leu Gln Gln Met Glu Ala Thr Ser Ala Ala
155 160 165
Thr Glu Ala Gly Leu Glu Leu Val Asp Lys Gly Lys Ala Lys Glu
170 175 180
Leu Pro Gly Ser Gln Val Ile Phe Glu Gly Pro Thr Leu Gly Gln
185 190 195
Gln Glu Asp Gln Glu Arg Lys Arg Leu Cys Lys Ala Met Thr Leu
200 205 210
Cys Ile Cys Tyr Ala Ala Ser Ile Gly Gly Thr Ala Thr Leu Thr
215 220 225
Gly Thr Gly Pro Asn Val Val Leu Leu Gly Gln Met Asn Glu Leu
230 235 240
Phe Pro Asp Ser Lys Asp Leu Val Asn Phe Ala Ser Trp Phe Ala
245 250 255
Phe Ala Phe Pro Asn Met Leu Val Met Leu Leu Phe Ala Trp Leu
260 265 270
Trp Leu Gln Phe Val Tyr Met Arg Phe Asn Phe Lys Lys Ser Trp
275 280 285
Gly Cys Gly Leu Glu Ser Lys Lys Asn Glu Lys Ala Ala Leu Lys
290 295 300
Val Leu Gln Glu Glu Tyr Arg Lys Leu Gly Pro Leu Ser Phe Ala
305 310 315
Glu Ile Asn Val Leu Ile Cys Phe Phe Leu Leu Val Ile Leu Trp

```
                320                 325                 330
Phe Ser Arg Asp Pro Gly Phe Met Pro Gly Trp Leu Thr Val Ala
                335                 340                 345
Trp Val Glu Glu Arg Lys Thr Pro Phe Tyr Pro Pro Pro Leu Leu
                350                 355                 360
Asp Trp Lys Val Thr Gln Glu Lys Val Pro Trp Gly Ile Val Leu
                365                 370                 375
Leu Leu Gly Gly Gly Phe Ala Leu Ala Lys Gly Ser Glu Ala Ser
                380                 385                 390
Gly Leu Ser Val Trp Met Gly Lys Gln Met Glu Pro Leu His Ala
                395                 400                 405
Val Pro Pro Ala Ala Ile Thr Leu Ile Leu Ser Leu Leu Val Ala
                410                 415                 420
Val Phe Thr Glu Cys Thr Ser Asn Val Ala Thr Thr Thr Leu Phe
                425                 430                 435
Leu Pro Ile Phe Ala Ser Met Ser Arg Ser Ile Gly Leu Asn Pro
                440                 445                 450
Leu Tyr Ile Met Leu Pro Cys Thr Leu Ser Ala Ser Phe Ala Phe
                455                 460                 465
Met Leu Pro Val Ala Thr Pro Pro Asn Ala Ile Val Phe Thr Tyr
                470                 475                 480
Gly His Leu Lys Val Ala Asp Met Val Lys Thr Gly Val Ile Met
                485                 490                 495
Asn Ile Ile Gly Val Phe Cys Val Phe Leu Ala Val Asn Thr Trp
                500                 505                 510
Gly Arg Ala Ile Phe Asp Leu Asp His Phe Pro Asp Trp Ala Asn
                515                 520                 525
Val Thr His Ile Glu Thr
                530
```

<210> 9
<211> 510
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500007CD1

```
<400> 9
Met Asp Ser Arg Val Ser Gly Thr Thr Ser Asn Gly Glu Thr Lys
  1               5                  10                  15
Pro Val Tyr Pro Val Met Glu Lys Lys Glu Glu Asp Gly Thr Leu
                20                  25                  30
Glu Arg Gly His Trp Asn Asn Lys Met Glu Phe Val Leu Ser Val
                35                  40                  45
Ala Gly Glu Ile Ile Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr
                50                  55                  60
Leu Cys Tyr Lys Asn Gly Gly Glu His Cys Met Glu Phe Gln Lys
                65                  70                  75
Thr Asn Gly Ser Leu Asn Gly Thr Ser Glu Asn Ala Thr Ser Pro
                80                  85                  90
Val Ile Glu Phe Trp Glu Arg Arg Val Leu Lys Ile Ser Asp Gly
                95                 100                 105
Ile Gln His Leu Gly Ala Leu Arg Trp Glu Leu Ala Leu Cys Leu
                110                 115                 120
```

```
Leu Leu Ala Trp Val Ile Cys Tyr Phe Cys Ile Trp Lys Gly Val
                125                 130                 135
Lys Ser Thr Gly Lys Val Val Tyr Phe Thr Ala Thr Phe Pro Tyr
                140                 145                 150
Leu Met Leu Val Val Leu Leu Ile Arg Gly Val Thr Leu Pro Gly
                155                 160                 165
Ala Ala Gln Gly Ile Gln Phe Tyr Leu Tyr Pro Asn Leu Thr Arg
                170                 175                 180
Leu Trp Asp Pro Gln Val Trp Met Asp Ala Gly Thr Gln Ile Phe
                185                 190                 195
Phe Ser Phe Ala Ile Cys Leu Gly Cys Leu Thr Ala Leu Gly Ser
                200                 205                 210
Tyr Asn Lys Tyr His Asn Asn Cys Tyr Arg Asp Cys Ile Ala Leu
                215                 220                 225
Cys Phe Leu Asn Ser Gly Thr Ser Phe Val Ala Gly Phe Ala Ile
                230                 235                 240
Phe Ser Ile Leu Gly Phe Met Ser Gln Glu Gln Gly Val Pro Ile
                245                 250                 255
Ser Glu Val Ala Glu Ser Gly Pro Gly Leu Ala Phe Ile Ala Tyr
                260                 265                 270
Pro Arg Ala Val Val Met Leu Pro Phe Ser Pro Leu Trp Ala Cys
                275                 280                 285
Cys Phe Phe Phe Met Val Val Leu Leu Gly Leu Asp Ser Gln Phe
                290                 295                 300
Val Cys Val Glu Ser Leu Val Thr Ala Leu Val Asp Met Tyr Pro
                305                 310                 315
His Val Phe Arg Lys Lys Asn Arg Arg Glu Val Leu Ile Leu Gly
                320                 325                 330
Val Ser Val Val Ser Phe Leu Val Gly Leu Ile Met Leu Thr Glu
                335                 340                 345
Gly Gly Met Tyr Val Phe Gln Leu Phe Asp Tyr Tyr Ala Ala Ser
                350                 355                 360
Gly Met Cys Leu Leu Phe Val Ala Ile Phe Glu Ser Leu Cys Val
                365                 370                 375
Ala Trp Val Tyr Gly Ala Lys Arg Phe Tyr Asp Asn Ile Glu Asp
                380                 385                 390
Met Ile Gly Tyr Arg Pro Trp Pro Leu Ile Lys Tyr Cys Trp Leu
                395                 400                 405
Phe Leu Thr Pro Ala Val Cys Thr Ala Thr Phe Leu Phe Ser Leu
                410                 415                 420
Ile Lys Tyr Thr Pro Leu Thr Tyr Asn Lys Lys Tyr Thr Tyr Pro
                425                 430                 435
Trp Trp Gly Asp Ala Leu Gly Trp Leu Leu Ala Leu Ser Ser Met
                440                 445                 450
Val Cys Ile Pro Ala Trp Ser Leu Tyr Arg Leu Gly Thr Leu Lys
                455                 460                 465
Gly Pro Phe Arg Glu Arg Ile Arg Gln Leu Met Cys Pro Ala Glu
                470                 475                 480
Asp Leu Pro Gln Arg Asn Pro Ala Gly Pro Ser Ala Pro Ala Thr
                485                 490                 495
Pro Arg Thr Ser Leu Leu Arg Leu Thr Glu Leu Glu Ser His Cys
                500                 505                 510
```

<210> 10
<211> 894

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500025CD1

<400> 10

```
Met Ser Val Arg Arg Gly His Val Ala Pro Gln Asn Thr Tyr Leu
  1               5                  10                  15
Asp Thr Ile Ile Arg Lys Phe Glu Gly Gln Ser Arg Lys Phe Leu
             20                  25                  30
Ile Ala Asn Ala Gln Met Glu Asn Cys Ala Ile Ile Tyr Cys Asn
             35                  40                  45
Asp Gly Phe Cys Glu Leu Phe Gly Tyr Ser Arg Val Glu Val Met
             50                  55                  60
Gln Gln Pro Cys Thr Cys Asp Phe Leu Thr Gly Pro Asn Thr Pro
             65                  70                  75
Ser Ser Ala Val Ser Arg Leu Ala Gln Ala Leu Leu Gly Ala Glu
             80                  85                  90
Glu Cys Lys Val Asp Ile Leu Tyr Tyr Arg Lys Asp Ala Ser Ser
             95                 100                 105
Phe Arg Cys Leu Val Asp Val Val Pro Val Lys Asn Glu Asp Gly
            110                 115                 120
Ala Val Ile Met Ser Ile Leu Asn Phe Glu Asp Leu Ala Gln Leu
            125                 130                 135
Leu Ala Lys Cys Ser Ser Arg Ser Leu Ser Gln Arg Leu Leu Ser
            140                 145                 150
Gln Ser Phe Leu Gly Ser Glu Gly Ser His Gly Arg Pro Gly Gly
            155                 160                 165
Pro Gly Pro Gly Thr Gly Arg Gly Lys Tyr Arg Thr Ile Ser Gln
            170                 175                 180
Ile Pro Gln Phe Thr Leu Asn Phe Val Glu Phe Asn Leu Glu Lys
            185                 190                 195
His Arg Ser Ser Ser Thr Thr Glu Ile Glu Ile Ile Ala Pro His
            200                 205                 210
Lys Val Val Glu Arg Thr Gln Asn Val Thr Glu Lys Val Thr Gln
            215                 220                 225
Val Leu Ser Leu Gly Ala Asp Val Leu Pro Glu Tyr Lys Leu Gln
            230                 235                 240
Ala Pro Arg Ile His Arg Trp Thr Ile Leu His Tyr Ser Pro Phe
            245                 250                 255
Lys Ala Val Trp Asp Trp Leu Ile Pro Leu Leu Val Ile Tyr Thr
            260                 265                 270
Ala Val Phe Thr Pro Tyr Ser Ala Ala Phe Leu Leu Ser Asp Gln
            275                 280                 285
Asp Glu Ser Arg Arg Gly Ala Cys Ser Tyr Thr Cys Ser Pro Leu
            290                 295                 300
Thr Val Val Asp Leu Ile Val Asp Ile Met Phe Val Val Asp Ile
            305                 310                 315
Val Ile Asn Phe Arg Thr Thr Tyr Val Asn Thr Asn Asp Glu Val
            320                 325                 330
Val Ser His Pro Arg Arg Ile Ala Val His Tyr Phe Lys Gly Trp
            335                 340                 345
Phe Leu Ile Asp Met Val Ala Ala Ile Pro Phe Asp Leu Leu Ile
            350                 355                 360
```

```
Phe Arg Thr Gly Ser Asp Glu Thr Thr Thr Leu Ile Gly Leu Leu
                365                 370                 375
Lys Thr Ala Arg Leu Leu Arg Leu Val Arg Val Ala Arg Lys Leu
                380                 385                 390
Asp Arg Tyr Ser Glu Tyr Gly Ala Ala Val Leu Phe Leu Leu Met
                395                 400                 405
Cys Thr Phe Ala Leu Ile Ala His Trp Leu Ala Cys Ile Cys Ser
                410                 415                 420
Leu Thr Ser Val Gly Phe Gly Asn Val Ser Pro Asn Thr Asn Ser
                425                 430                 435
Glu Lys Val Phe Ser Ile Cys Val Met Leu Ile Gly Ser Leu Met
                440                 445                 450
Tyr Ala Ser Ile Phe Gly Asn Val Ser Ala Ile Ile Gln Arg Leu
                455                 460                 465
Tyr Ser Gly Thr Ala Arg Tyr His Thr Gln Met Leu Arg Val Lys
                470                 475                 480
Glu Phe Ile Arg Phe His Gln Ile Pro Asn Pro Leu Arg Gln Arg
                485                 490                 495
Leu Glu Glu Tyr Phe Gln His Ala Trp Ser Tyr Thr Asn Gly Ile
                500                 505                 510
Asp Met Asn Ala Val Leu Lys Gly Phe Pro Glu Cys Leu Gln Ala
                515                 520                 525
Asp Ile Cys Leu His Leu His Arg Ala Leu Leu Gln His Cys Pro
                530                 535                 540
Ala Phe Ser Gly Ala Gly Lys Gly Cys Leu Arg Ala Leu Ala Val
                545                 550                 555
Lys Phe Lys Thr Thr His Ala Pro Pro Gly Asp Thr Leu Val His
                560                 565                 570
Leu Gly Asp Val Leu Ser Thr Leu Tyr Phe Ile Ser Arg Gly Ser
                575                 580                 585
Ile Glu Ile Leu Arg Asp Asp Val Val Val Ala Ile Leu Gly Lys
                590                 595                 600
Asn Asp Ile Phe Gly Glu Pro Val Ser Leu His Ala Gln Pro Gly
                605                 610                 615
Lys Ser Ser Ala Asp Val Arg Ala Leu Thr Tyr Cys Asp Leu His
                620                 625                 630
Lys Ile Gln Arg Ala Asp Leu Leu Glu Val Leu Asp Met Tyr Pro
                635                 640                 645
Ala Phe Ala Glu Ser Phe Trp Ser Lys Leu Glu Val Thr Phe Asn
                650                 655                 660
Leu Arg Asp Ala Pro Gly Ser Gln Asp His Gln Gly Phe Phe Leu
                665                 670                 675
Ser Asp Asn Gln Ser Asp Ala Ala Pro Pro Leu Ser Ile Ser Asp
                680                 685                 690
Ala Ser Gly Leu Trp Pro Glu Leu Leu Gln Glu Met Pro Pro Arg
                695                 700                 705
His Ser Pro Gln Ser Pro Gln Glu Asp Pro Asp Cys Trp Pro Leu
                710                 715                 720
Lys Leu Gly Ser Arg Leu Glu Gln Leu Gln Ala Gln Met Asn Arg
                725                 730                 735
Leu Glu Ser Arg Val Ser Ser Asp Leu Ser Arg Ile Leu Gln Leu
                740                 745                 750
Leu Gln Lys Pro Met Pro Gln Gly His Ala Ser Tyr Ile Leu Glu
                755                 760                 765
Ala Pro Ala Ser Asn Asp Leu Ala Leu Val Pro Ile Ala Ser Glu
                770                 775                 780
```

```
Thr Thr Ser Pro Gly Pro Arg Leu Pro Gln Gly Phe Leu Pro Pro
                785                 790                 795
Ala Gln Thr Pro Ser Tyr Gly Asp Leu Asp Asp Cys Ser Pro Lys
                800                 805                 810
His Arg Asn Ser Ser Pro Arg Met Pro His Leu Ala Val Ala Met
                815                 820                 825
Asp Lys Thr Leu Ala Pro Ser Ser Glu Gln Glu Gln Pro Glu Gly
                830                 835                 840
Leu Trp Pro Pro Leu Ala Ser Pro Leu His Pro Leu Glu Val Gln
                845                 850                 855
Gly Leu Ile Cys Gly Pro Cys Phe Ser Ser Leu Pro Glu His Leu
                860                 865                 870
Gly Ser Val Pro Lys Gln Leu Asp Phe Gln Arg His Gly Ser Asp
                875                 880                 885
Pro Gly Phe Ala Gly Ser Trp Gly His
                890
```

<210> 11
<211> 788
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7502736CD1

<400> 11

```
Met Ala Gly Gln Ala Asp Gln Gly Gln Ala Gln Ala Gly Ala Ser
  1               5                  10                  15
Thr Gly Pro Ser Ala Arg Ser His Ser Ser Arg Ser Thr Ser Trp
                 20                  25                  30
Ser Ser Thr Trp Arg Ser Thr Ala Pro Ala Pro Pro Arg Arg Leu
                 35                  40                  45
Arg Ser Ser Arg Pro Ile Arg Trp Trp Ser Gly His Arg Thr Ser
                 50                  55                  60
Leu Arg Arg Ser Pro Arg Ser Cys Pro Trp Ala Arg Met Cys Cys
                 65                  70                  75
Arg Ser Thr Ser Cys Arg Arg Arg Ala Ser Thr Ala Gly Pro Ser
                 80                  85                  90
Cys Thr Thr Ala Pro Phe Lys Ala Val Trp Asp Trp Leu Ile Leu
                 95                 100                 105
Leu Leu Val Ile Tyr Thr Ala Val Phe Thr Pro Tyr Ser Ala Ala
                110                 115                 120
Phe Leu Leu Ser Asp Gln Asp Glu Ser Arg Arg Gly Ala Cys Ser
                125                 130                 135
Tyr Thr Cys Ser Pro Leu Thr Val Val Asp Leu Ile Val Asp Ile
                140                 145                 150
Met Phe Val Val Asp Ile Val Ile Asn Phe Arg Thr Thr Tyr Val
                155                 160                 165
Asn Thr Asn Asp Glu Val Val Ser His Pro Arg Arg Ile Ala Val
                170                 175                 180
His Tyr Phe Lys Gly Trp Phe Leu Ile Asp Met Val Ala Ala Ile
                185                 190                 195
Pro Phe Asp Leu Leu Ile Phe Arg Thr Gly Ser Asp Glu Thr Thr
                200                 205                 210
Thr Leu Ile Gly Leu Leu Lys Thr Ala Arg Leu Leu Arg Leu Val
```

215 220 225
Arg Val Ala Arg Lys Leu Asp Arg Tyr Ser Glu Tyr Gly Ala Ala
230 235 240
Val Leu Phe Leu Leu Met Cys Thr Phe Ala Leu Ile Ala His Trp
245 250 255
Leu Ala Cys Ile Trp Tyr Ala Ile Gly Asn Val Glu Arg Pro Tyr
260 265 270
Leu Glu His Lys Ile Gly Trp Leu Asp Ser Leu Gly Val Gln Leu
275 280 285
Gly Lys Arg Tyr Asn Gly Ser Asp Pro Ala Ser Gly Pro Ser Val
290 295 300
Gln Asp Lys Tyr Val Thr Ala Leu Tyr Phe Thr Phe Ser Ser Leu
305 310 315
Thr Ser Val Gly Phe Gly Asn Val Ser Pro Asn Thr Asn Ser Glu
320 325 330
Lys Val Phe Ser Ile Cys Val Met Leu Ile Gly Ser Leu Met Tyr
335 340 345
Ala Ser Ile Phe Gly Asn Val Ser Ala Ile Ile Gln Arg Leu Tyr
350 355 360
Ser Gly Thr Ala Arg Tyr His Thr Gln Met Leu Arg Val Lys Glu
365 370 375
Phe Ile Arg Phe His Gln Ile Pro Asn Pro Leu Arg Gln Arg Leu
380 385 390
Glu Glu Tyr Phe Gln His Ala Trp Ser Tyr Thr Asn Gly Ile Asp
395 400 405
Met Asn Ala Val Leu Lys Gly Phe Pro Glu Cys Leu Gln Ala Asp
410 415 420
Ile Cys Leu His Leu His Arg Ala Leu Leu Gln His Cys Pro Ala
425 430 435
Phe Ser Gly Ala Gly Lys Gly Cys Leu Arg Ala Leu Ala Val Lys
440 445 450
Phe Lys Thr Thr His Ala Pro Pro Gly Asp Thr Leu Val His Leu
455 460 465
Gly Asp Val Leu Ser Thr Leu Tyr Phe Ile Ser Arg Gly Ser Ile
470 475 480
Glu Ile Leu Arg Asp Asp Val Val Val Ala Ile Leu Gly Lys Asn
485 490 495
Asp Ile Phe Gly Glu Pro Val Ser Leu His Ala Gln Pro Gly Lys
500 505 510
Ser Ser Ala Asp Val Arg Ala Leu Thr Tyr Cys Asp Leu His Lys
515 520 525
Ile Gln Arg Ala Asp Leu Leu Glu Val Leu Asp Met Tyr Pro Ala
530 535 540
Phe Ala Glu Ser Phe Trp Ser Lys Leu Glu Val Thr Phe Asn Leu
545 550 555
Arg Asp Ala Pro Gly Ser Gln Asp His Gln Gly Phe Phe Leu Ser
560 565 570
Asp Asn Gln Ser Asp Ala Ala Pro Pro Leu Ser Ile Ser Asp Ala
575 580 585
Ser Gly Leu Trp Pro Glu Leu Leu Gln Glu Met Pro Pro Arg His
590 595 600
Ser Pro Gln Ser Pro Gln Glu Asp Pro Asp Cys Trp Pro Leu Lys
605 610 615
Leu Gly Ser Arg Leu Glu Gln Leu Gln Ala Gln Met Asn Arg Leu
620 625 630
Glu Ser Arg Val Ser Ser Asp Leu Ser Arg Ile Leu Gln Leu Leu

635 640 645
Gln Lys Pro Met Pro Gln Gly His Ala Ser Tyr Ile Leu Glu Ala
650 655 660
Pro Ala Ser Asn Asp Leu Ala Leu Val Pro Ile Ala Ser Glu Thr
665 670 675
Thr Ser Pro Gly Pro Arg Leu Pro Gln Gly Phe Leu Pro Pro Ala
680 685 690
Gln Thr Pro Ser Tyr Gly Asp Leu Asp Asp Cys Ser Pro Lys His
695 700 705
Arg Asn Ser Ser Pro Arg Met Pro His Leu Ala Val Ala Met Asp
710 715 720
Lys Thr Leu Ala Pro Ser Ser Glu Gln Glu Gln Pro Glu Gly Leu
725 730 735
Trp Pro Pro Leu Ala Ser Pro Leu His Pro Leu Glu Val Gln Gly
740 745 750
Leu Ile Cys Gly Pro Cys Phe Ser Ser Leu Pro Glu His Leu Gly
755 760 765
Ser Val Pro Lys Gln Leu Asp Phe Gln Arg His Gly Ser Asp Pro
770 775 780
Gly Phe Ala Gly Ser Trp Gly His
785

<210> 12
<211> 501
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503570CD1

<400> 12
Met Ser Gln Asp Thr Glu Val Asp Met Lys Glu Val Glu Leu Asn
1 5 10 15
Glu Leu Glu Pro Glu Lys Gln Pro Met Asn Ala Ala Ser Gly Ala
20 25 30
Ala Met Ser Leu Ala Gly Ala Glu Lys Asn Gly Leu Val Lys Ile
35 40 45
Lys Val Ala Glu Asp Glu Ala Glu Ala Ala Ala Ala Ala Lys Phe
50 55 60
Thr Gly Leu Ser Lys Glu Glu Leu Leu Lys Val Ala Gly Ser Pro
65 70 75
Gly Trp Val Arg Thr Arg Trp Ala Leu Leu Leu Leu Phe Trp Leu
80 85 90
Gly Trp Leu Gly Met Leu Ala Gly Ala Val Val Ile Ile Val Arg
95 100 105
Ala Pro Arg Cys Arg Glu Leu Pro Ala Gln Lys Trp Trp His Thr
110 115 120
Gly Ala Leu Tyr Arg Ile Gly Asp Leu Gln Ala Phe Gln Gly His
125 130 135
Gly Ala Gly Asn Leu Ala Gly Leu Lys Gly Arg Leu Asp Tyr Leu
140 145 150
Ser Ser Leu Lys Val Lys Gly Leu Val Leu Gly Pro Ile His Lys
155 160 165
Asn Gln Lys Asp Asp Val Ala Gln Thr Asp Leu Leu Gln Ile Asp
170 175 180

```
Pro Asn Phe Gly Ser Lys Glu Asp Phe Asp Ser Leu Leu Gln Ser
                185                 190                 195
Ala Lys Lys Lys Ser Ile Arg Val Ile Leu Asp Leu Thr Pro Asn
                200                 205                 210
Tyr Arg Gly Glu Asn Ser Trp Phe Ser Thr Gln Val Asp Thr Val
                215                 220                 225
Ala Thr Lys Val Lys Asp Ala Leu Glu Phe Trp Leu Gln Ala Gly
                230                 235                 240
Val Asp Gly Phe Gln Val Arg Asp Ile Glu Asn Leu Lys Asp Ala
                245                 250                 255
Ser Ser Phe Leu Ala Glu Trp Gln Asn Ile Thr Lys Gly Phe Ser
                260                 265                 270
Glu Asp Arg Leu Leu Ile Ala Gly Thr Asn Ser Ser Asp Leu Gln
                275                 280                 285
Gln Ile Leu Ser Leu Leu Glu Ser Asn Lys Asp Leu Leu Leu Thr
                290                 295                 300
Ser Ser Tyr Leu Ser Asp Ser Gly Ser Thr Gly Glu His Thr Lys
                305                 310                 315
Ser Leu Val Thr Gln Tyr Leu Asn Ala Thr Gly Asn Arg Trp Cys
                320                 325                 330
Ser Trp Ser Leu Ser Gln Ala Arg Leu Leu Thr Ser Phe Leu Pro
                335                 340                 345
Ala Gln Leu Leu Arg Leu Tyr Gln Leu Met Leu Phe Thr Leu Pro
                350                 355                 360
Gly Thr Pro Val Phe Ser Tyr Gly Asp Glu Ile Gly Leu Asp Ala
                365                 370                 375
Ala Ala Leu Pro Gly Gln Gly Gln Ser Glu Asp Pro Gly Ser Leu
                380                 385                 390
Leu Ser Leu Phe Arg Arg Leu Ser Asp Gln Arg Ser Lys Glu Arg
                395                 400                 405
Ser Leu Leu His Gly Asp Phe His Ala Phe Ser Ala Gly Pro Gly
                410                 415                 420
Leu Phe Ser Tyr Ile Arg His Trp Asp Gln Asn Glu Arg Phe Leu
                425                 430                 435
Val Val Leu Asn Phe Gly Asp Val Gly Leu Ser Ala Gly Leu Gln
                440                 445                 450
Ala Ser Asp Leu Pro Ala Ser Ala Ser Leu Pro Ala Lys Ala Asp
                455                 460                 465
Leu Leu Leu Ser Thr Gln Pro Gly Arg Glu Glu Gly Ser Pro Leu
                470                 475                 480
Glu Leu Glu Arg Leu Lys Leu Glu Pro His Glu Gly Leu Leu Leu
                485                 490                 495
Arg Phe Pro Tyr Ala Ala
                500
```

<210> 13
<211> 721
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7504008CD1

<400> 13
Met Gly Leu Ala Asp Ala Ser Gly Pro Arg Asp Thr Gln Ala Leu

```
  1               5                  10                  15
Leu Ser Ala Thr Gln Ala Met Asp Leu Arg Arg Arg Asp Tyr His
             20                  25                  30
Met Glu Arg Pro Leu Leu Asn Gln Glu His Leu Glu Glu Leu Gly
             35                  40                  45
Arg Trp Gly Ser Ala Pro Arg Thr His Gln Trp Arg Thr Trp Leu
             50                  55                  60
Gln Cys Ser Arg Ala Arg Ala Tyr Ala Leu Leu Leu Gln His Leu
             65                  70                  75
Pro Val Leu Val Trp Leu Pro Arg Tyr Pro Val Arg Asp Trp Leu
             80                  85                  90
Leu Gly Asp Leu Leu Ser Gly Leu Ser Val Ala Ile Met Gln Leu
             95                 100                 105
Pro Gln Gly Leu Ala Tyr Ala Leu Leu Ala Gly Leu Pro Pro Val
            110                 115                 120
Phe Gly Leu Tyr Ser Ser Phe Tyr Pro Val Phe Ile Tyr Phe Leu
            125                 130                 135
Phe Gly Thr Ser Arg His Ile Ser Val Gly Thr Phe Ala Val Met
            140                 145                 150
Ser Val Met Val Gly Ser Val Thr Glu Ser Leu Ala Pro Gln Ala
            155                 160                 165
Leu Asn Asp Ser Met Ile Asn Glu Thr Ala Arg Asp Ala Ala Arg
            170                 175                 180
Val Gln Val Ala Ser Thr Leu Ser Val Leu Val Gly Leu Phe Gln
            185                 190                 195
Val Gly Leu Gly Leu Ile His Phe Gly Phe Val Val Thr Tyr Leu
            200                 205                 210
Ser Glu Pro Leu Val Arg Gly Tyr Thr Thr Ala Ala Ala Val Gln
            215                 220                 225
Val Phe Val Ser Gln Leu Lys Tyr Val Phe Gly Leu His Leu Ser
            230                 235                 240
Ser His Ser Gly Pro Leu Ser Leu Ile Tyr Thr Val Leu Glu Val
            245                 250                 255
Cys Trp Lys Leu Pro Gln Ser Lys Leu Ile Gly Ala Thr Gly Ile
            260                 265                 270
Ser Tyr Gly Met Gly Leu Lys His Arg Phe Glu Val Asp Val Val
            275                 280                 285
Gly Asn Ile Pro Ala Gly Leu Val Pro Pro Val Ala Pro Asn Thr
            290                 295                 300
Gln Leu Phe Ser Lys Leu Val Gly Ser Ala Phe Thr Ile Ala Val
            305                 310                 315
Val Gly Phe Ala Ile Ala Ile Ser Leu Gly Lys Ile Phe Ala Leu
            320                 325                 330
Arg His Gly Tyr Arg Val Asp Ser Asn Gln Glu Leu Val Ala Leu
            335                 340                 345
Gly Leu Ser Asn Leu Ile Gly Gly Ile Phe Gln Cys Phe Pro Val
            350                 355                 360
Ser Cys Ser Met Ser Arg Ser Leu Val Gln Glu Ser Thr Gly Gly
            365                 370                 375
Asn Ser Gln Val Ala Gly Ala Ile Ser Ser Leu Phe Ile Leu Leu
            380                 385                 390
Ile Ile Val Lys Leu Gly Glu Leu Phe His Asp Leu Pro Lys Ala
            395                 400                 405
Val Leu Ala Ala Ile Ile Ile Val Asn Leu Lys Gly Met Leu Arg
            410                 415                 420
Gln Leu Ser Asp Met Arg Ser Leu Trp Lys Ala Asn Arg Ala Asp
```

```
                425                 430                 435
Leu Leu Ile Trp Leu Val Thr Phe Thr Ala Thr Ile Leu Leu Asn
                440                 445                 450
Leu Asp Leu Gly Leu Val Val Ala Val Ile Phe Ser Leu Leu Leu
                455                 460                 465
Val Val Val Arg Thr Gln Met Pro His Tyr Ser Val Leu Gly Gln
                470                 475                 480
Val Pro Asp Thr Asp Ile Tyr Arg Asp Val Ala Glu Tyr Ser Glu
                485                 490                 495
Ala Lys Glu Val Arg Gly Val Lys Val Phe Arg Ser Ser Ala Thr
                500                 505                 510
Val Tyr Phe Ala Asn Ala Glu Phe Tyr Ser Asp Ala Leu Lys Gln
                515                 520                 525
Arg Cys Gly Val Asp Val Asp Phe Leu Ile Ser Gln Lys Lys Lys
                530                 535                 540
Leu Leu Lys Lys Gln Glu Gln Leu Lys Leu Lys Gln Leu Gln Lys
                545                 550                 555
Glu Glu Lys Leu Arg Lys Gln Ala Ala Ser Pro Lys Gly Ala Ser
                560                 565                 570
Val Ser Ile Asn Val Asn Thr Ser Leu Glu Asp Met Arg Ser Asn
                575                 580                 585
Asn Val Glu Asp Cys Lys Met Met Gln Val Ser Ser Gly Asp Lys
                590                 595                 600
Met Glu Asp Ala Thr Ala Asn Gly Gln Glu Asp Ser Lys Ala Pro
                605                 610                 615
Asp Gly Ser Thr Leu Lys Ala Leu Gly Leu Pro Gln Pro Asp Phe
                620                 625                 630
His Ser Leu Ile Leu Asp Leu Gly Ala Leu Ser Phe Val Asp Thr
                635                 640                 645
Val Cys Leu Lys Ser Leu Lys Asn Ile Phe His Asp Phe Arg Glu
                650                 655                 660
Ile Glu Val Glu Val Tyr Met Ala Ala Cys His Ser Pro Val Val
                665                 670                 675
Ser Gln Leu Glu Ala Gly His Phe Phe Asp Ala Ser Ile Thr Lys
                680                 685                 690
Lys His Leu Phe Ala Ser Val His Asp Ala Val Thr Phe Ala Leu
                695                 700                 705
Gln His Pro Arg Pro Val Pro Asp Ser Pro Val Ser Val Thr Arg
                710                 715                 720
Leu
```

<210> 14
<211> 1226
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503559CD1

```
<400> 14
Met Asp Arg Glu Glu Arg Lys Thr Ile Asn Gln Gly Gln Glu Asp
  1               5                  10                  15
Glu Met Glu Ile Tyr Gly Tyr Asn Leu Ser Arg Trp Lys Leu Ala
                 20                  25                  30
```

```
Ile Val Ser Leu Gly Val Ile Cys Ser Gly Gly Phe Leu Leu Leu
                35                  40                  45
Leu Leu Tyr Trp Met Pro Glu Trp Arg Val Lys Ala Thr Cys Val
                50                  55                  60
Arg Ala Ala Ile Lys Asp Cys Glu Val Val Leu Leu Arg Thr Thr
                65                  70                  75
Asp Glu Phe Lys Met Trp Phe Cys Ala Lys Ile Arg Val Leu Ser
                80                  85                  90
Leu Glu Thr Tyr Pro Val Ser Ser Pro Lys Ser Met Ser Asn Lys
                95                 100                 105
Leu Ser Asn Gly His Ala Val Cys Leu Ile Glu Asn Pro Thr Glu
               110                 115                 120
Glu Asn Arg His Arg Ile Ser Lys Tyr Ser Gln Thr Glu Ser Gln
               125                 130                 135
Gln Ile Arg Tyr Phe Thr His His Ser Val Lys Tyr Phe Trp Asn
               140                 145                 150
Asp Thr Ile His Asn Phe Asp Phe Leu Lys Gly Leu Asp Glu Gly
               155                 160                 165
Val Ser Cys Thr Ser Ile Tyr Glu Lys His Ser Ala Gly Leu Thr
               170                 175                 180
Lys Gly Met His Ala Tyr Arg Lys Leu Leu Tyr Gly Val Asn Glu
               185                 190                 195
Ile Ala Val Lys Val Pro Ser Val Phe Lys Leu Leu Ile Lys Glu
               200                 205                 210
Val Leu Asn Pro Phe Tyr Ile Phe Gln Leu Phe Ser Val Ile Leu
               215                 220                 225
Trp Ser Thr Asp Glu Tyr Tyr Tyr Tyr Ala Leu Ala Ile Val Val
               230                 235                 240
Met Ser Ile Val Ser Ile Val Ser Ser Leu Tyr Ser Ile Arg Lys
               245                 250                 255
Gln Tyr Val Met Leu His Asp Met Val Ala Thr His Ser Thr Val
               260                 265                 270
Arg Val Ser Val Cys Arg Val Asn Glu Glu Ile Glu Glu Ile Phe
               275                 280                 285
Ser Thr Asp Leu Val Pro Gly Asp Val Met Val Ile Pro Leu Asn
               290                 295                 300
Gly Thr Ile Met Pro Cys Asp Ala Val Leu Ile Asn Gly Thr Cys
               305                 310                 315
Ile Val Asn Glu Ser Met Leu Thr Gly Glu Ser Val Pro Val Thr
               320                 325                 330
Lys Thr Asn Leu Pro Asn Pro Ser Val Asp Val Lys Gly Ile Gly
               335                 340                 345
Asp Glu Leu Tyr Asn Pro Glu Thr His Lys Arg His Thr Leu Phe
               350                 355                 360
Cys Gly Thr Thr Val Ile Gln Thr Arg Phe Tyr Thr Gly Glu Leu
               365                 370                 375
Val Lys Ala Ile Val Val Arg Thr Gly Phe Ser Thr Ser Lys Gly
               380                 385                 390
Gln Leu Val Arg Ser Ile Leu Tyr Pro Lys Pro Thr Asp Phe Lys
               395                 400                 405
Leu Tyr Arg Asp Ala Tyr Leu Phe Leu Leu Cys Leu Val Ala Val
               410                 415                 420
Ala Gly Ile Gly Phe Ile Tyr Thr Ile Ile Asn Ser Ile Leu Asn
               425                 430                 435
Glu Val Gln Val Gly Val Ile Ile Ile Glu Ser Leu Asp Ile Ile
               440                 445                 450
```

```
Thr Ile Thr Val Pro Pro Ala Leu Pro Ala Ala Met Thr Ala Gly
                455                 460                 465
Ile Val Tyr Ala Gln Arg Arg Leu Lys Lys Ile Gly Ile Phe Cys
                470                 475                 480
Ile Ser Pro Gln Arg Ile Asn Ile Cys Gly Gln Leu Asn Leu Val
                485                 490                 495
Cys Phe Asp Lys Thr Gly Thr Leu Thr Glu Asp Gly Leu Asp Leu
                500                 505                 510
Trp Gly Ile Gln Arg Val Glu Asn Ala Arg Phe Leu Ser Pro Glu
                515                 520                 525
Glu Asn Val Cys Asn Glu Met Leu Val Lys Ser Gln Phe Val Ala
                530                 535                 540
Cys Met Ala Thr Cys His Ser Leu Thr Lys Ile Glu Gly Val Leu
                545                 550                 555
Ser Gly Asp Pro Leu Asp Leu Lys Met Phe Glu Ala Ile Gly Trp
                560                 565                 570
Ile Leu Glu Glu Ala Thr Glu Glu Glu Thr Ala Leu His Asn Arg
                575                 580                 585
Ile Met Pro Thr Val Val Arg Pro Pro Lys Gln Leu Leu Pro Glu
                590                 595                 600
Ser Thr Pro Ala Gly Asn Gln Glu Met Glu Leu Phe Glu Leu Pro
                605                 610                 615
Ala Thr Tyr Glu Ile Gly Ile Val Arg Gln Phe Pro Phe Ser Ser
                620                 625                 630
Ala Leu Gln Arg Met Ser Val Val Ala Arg Val Leu Gly Asp Arg
                635                 640                 645
Lys Met Asp Ala Tyr Met Lys Gly Ala Pro Glu Ala Ile Ala Gly
                650                 655                 660
Leu Cys Lys Pro Glu Thr Val Pro Val Asp Phe Gln Asn Val Leu
                665                 670                 675
Glu Asp Phe Thr Lys Gln Gly Phe Arg Val Ile Ala Leu Ala His
                680                 685                 690
Arg Lys Leu Glu Ser Lys Leu Thr Trp His Lys Val Gln Asn Ile
                695                 700                 705
Ser Arg Asp Ala Ile Glu Asn Asn Met Asp Phe Met Gly Leu Ile
                710                 715                 720
Ile Met Gln Asn Lys Leu Lys Gln Glu Thr Pro Ala Val Leu Glu
                725                 730                 735
Asp Leu His Lys Ala Asn Ile Arg Thr Val Met Val Thr Gly Asp
                740                 745                 750
Ser Met Leu Thr Ala Val Ser Val Ala Arg Asp Cys Gly Met Ile
                755                 760                 765
Leu Pro Gln Asp Lys Val Ile Ile Ala Glu Ala Leu Pro Pro Lys
                770                 775                 780
Asp Gly Lys Val Ala Lys Ile Asn Trp His Tyr Ala Asp Ser Leu
                785                 790                 795
Thr Gln Cys Ser His Pro Ser Ala Ile Asp Pro Glu Ala Ile Pro
                800                 805                 810
Val Lys Leu Val His Asp Ser Leu Glu Asp Leu Gln Met Thr Arg
                815                 820                 825
Tyr His Phe Ala Met Asn Gly Lys Ser Phe Ser Val Ile Leu Glu
                830                 835                 840
His Phe Gln Asp Leu Val Pro Lys Leu Met Leu His Gly Thr Val
                845                 850                 855
Phe Ala Arg Met Ala Pro Asp Gln Lys Thr Gln Leu Ile Glu Ala
                860                 865                 870
```

```
Leu Gln Asn Val Asp Tyr Phe Val Gly Met Cys Gly Asp Gly Ala
                875                 880                 885
Asn Asp Cys Gly Ala Leu Lys Arg Ala His Gly Gly Ile Ser Leu
                890                 895                 900
Ser Glu Leu Glu Ala Ser Val Ala Ser Pro Phe Thr Ser Lys Thr
                905                 910                 915
Pro Ser Ile Ser Cys Val Pro Asn Leu Ile Arg Glu Gly Arg Ala
                920                 925                 930
Ala Leu Ile Thr Ser Phe Cys Val Phe Lys Phe Met Ala Leu Tyr
                935                 940                 945
Ser Ile Ile Gln Tyr Phe Ser Val Thr Leu Leu Tyr Ser Ile Leu
                950                 955                 960
Ser Asn Leu Gly Asp Phe Gln Phe Leu Phe Ile Asp Leu Ala Ile
                965                 970                 975
Ile Leu Val Val Val Phe Thr Met Ser Leu Asn Pro Ala Trp Lys
                980                 985                 990
Glu Leu Val Ala Gln Arg Pro Pro Ser Gly Leu Ile Ser Gly Ala
                995                1000                1005
Leu Leu Phe Ser Val Leu Ser Gln Ile Ile Ile Cys Ile Gly Phe
               1010                1015                1020
Gln Ser Leu Gly Phe Phe Trp Val Lys Gln Gln Pro Trp Tyr Glu
               1025                1030                1035
Val Trp His Pro Lys Ser Asp Ala Cys Asn Thr Thr Gly Ser Gly
               1040                1045                1050
Phe Trp Asn Ser Ser His Val Asp Asn Glu Thr Glu Leu Asp Glu
               1055                1060                1065
His Asn Ile Gln Asn Tyr Glu Asn Thr Thr Val Phe Phe Ile Ser
               1070                1075                1080
Ser Phe Gln Tyr Leu Ile Val Ala Ile Ala Phe Ser Lys Gly Lys
               1085                1090                1095
Pro Phe Arg Gln Pro Cys Tyr Lys Asn Tyr Phe Phe Val Phe Ser
               1100                1105                1110
Val Ile Phe Leu Tyr Ile Phe Ile Leu Phe Ile Met Leu Tyr Pro
               1115                1120                1125
Val Ala Ser Val Asp Gln Val Leu Gln Ile Val Cys Val Pro Tyr
               1130                1135                1140
Gln Trp Arg Val Thr Met Leu Ile Ile Val Leu Val Asn Ala Phe
               1145                1150                1155
Val Ser Ile Thr Val Glu Glu Ser Val Asp Arg Trp Gly Lys Cys
               1160                1165                1170
Cys Leu Pro Trp Ala Leu Gly Cys Arg Lys Lys Thr Pro Lys Ala
               1175                1180                1185
Lys Tyr Met Tyr Leu Ala Gln Glu Leu Leu Val Asp Pro Glu Trp
               1190                1195                1200
Pro Pro Lys Pro Gln Thr Thr Thr Glu Ala Lys Ala Leu Val Lys
               1205                1210                1215
Glu Asn Gly Ser Cys Gln Ile Ile Thr Ile Thr
               1220                1225
```

<210> 15
<211> 638
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 6243872CD1

<400> 15

```
Met Phe Val Gly Val Ala Arg His Ser Gly Ser Gln Asp Glu Val
  1               5                  10                  15
Ser Arg Gly Val Glu Pro Leu Glu Ala Ala Arg Ala Gln Pro Ala
                 20                  25                  30
Lys Asp Arg Arg Ala Lys Gly Thr Pro Lys Ser Ser Lys Pro Gly
                 35                  40                  45
Lys Lys His Arg Tyr Leu Arg Leu Leu Pro Glu Ala Leu Ile Arg
                 50                  55                  60
Phe Gly Gly Phe Arg Lys Arg Lys Lys Ala Lys Ser Ser Val Ser
                 65                  70                  75
Lys Lys Pro Gly Glu Val Asp Asp Ser Leu Glu Gln Pro Cys Gly
                 80                  85                  90
Leu Gly Cys Leu Val Ser Thr Cys Cys Glu Cys Cys Asn Asn Ile
                 95                 100                 105
Arg Cys Phe Met Ile Phe Tyr Cys Ile Leu Leu Ile Cys Gln Gly
                110                 115                 120
Val Val Phe Gly Leu Ile Asp Val Ser Ile Gly Asp Phe Gln Lys
                125                 130                 135
Glu Tyr Gln Leu Lys Thr Ile Glu Lys Leu Ala Leu Glu Lys Ser
                140                 145                 150
Tyr Asp Ile Ser Ser Gly Leu Thr Val Gln Gly Ile Ala Gly Met
                155                 160                 165
Pro Leu Tyr Ile Leu Gly Ile Thr Phe Ile Asp Glu Asn Val Ala
                170                 175                 180
Thr His Ser Ala Gly Ile Tyr Leu Gly Ile Ala Glu Cys Thr Ser
                185                 190                 195
Met Ile Gly Tyr Ala Leu Gly Tyr Val Leu Gly Ala Pro Leu Val
                200                 205                 210
Lys Val Pro Glu Asn Thr Thr Ser Ala Thr Asn Thr Thr Val Asn
                215                 220                 225
Asn Gly Ser Pro Glu Trp Leu Trp Thr Trp Trp Ile Asn Phe Leu
                230                 235                 240
Phe Ala Ala Val Val Ala Trp Cys Thr Leu Ile Pro Leu Ser Cys
                245                 250                 255
Phe Pro Asn Asn Met Pro Gly Ser Thr Arg Ile Lys Ala Arg Lys
                260                 265                 270
Arg Lys Gln Leu His Phe Phe Asp Ser Arg Leu Lys Asp Leu Lys
                275                 280                 285
Leu Gly Ile Asn Ile Lys Asp Leu Cys Ala Ala Leu Trp Ile Leu
                290                 295                 300
Met Lys Asn Pro Val Leu Ile Cys Leu Ala Leu Ser Lys Ala Thr
                305                 310                 315
Glu Tyr Leu Val Ile Ile Gly Ala Ser Glu Phe Leu Pro Ile Tyr
                320                 325                 330
Leu Glu Asn Gln Phe Ile Leu Thr Pro Thr Val Ala Thr Thr Leu
                335                 340                 345
Ala Gly Leu Val Leu Ile Pro Gly Gly Ala Leu Gly Gln Leu Leu
                350                 355                 360
Gly Gly Val Ile Val Ser Thr Leu Glu Met Ser Cys Lys Ala Leu
                365                 370                 375
Met Arg Phe Ile Met Val Thr Ser Val Ile Ser Leu Ile Leu Leu
                380                 385                 390
Val Phe Ile Ile Phe Val Arg Cys Asn Pro Val Gln Phe Ala Gly
```

```
                395                 400                 405
Ile Asn Glu Asp Tyr Asp Gly Thr Gly Lys Leu Gly Asn Leu Thr
                410                 415                 420
Ala Pro Cys Asn Glu Lys Cys Arg Cys Ser Ser Ser Ile Tyr Ser
                425                 430                 435
Ser Ile Cys Gly Arg Asp Asp Ile Glu Tyr Phe Ser Pro Cys Phe
                440                 445                 450
Ala Gly Ile Val Ser Cys Leu Gln Tyr Ser Gln Met Tyr Tyr Asn
                455                 460                 465
Cys Ser Cys Ile Lys Glu Gly Leu Ile Thr Ala Asp Ala Glu Gly
                470                 475                 480
Asp Phe Ile Asp Ala Arg Pro Gly Lys Cys Asp Ala Lys Cys Tyr
                485                 490                 495
Lys Leu Pro Leu Phe Ile Ala Phe Ile Phe Ser Thr Leu Ile Phe
                500                 505                 510
Ser Gly Phe Ser Gly Val Pro Ile Val Leu Ala Met Thr Arg Val
                515                 520                 525
Val Pro Asp Lys Leu Arg Ser Leu Ala Leu Gly Val Ser Tyr Val
                530                 535                 540
Ile Leu Arg Ile Phe Gly Thr Ile Pro Gly Pro Ser Ile Phe Lys
                545                 550                 555
Met Ser Gly Glu Thr Ser Cys Ile Leu Arg Asp Val Asn Lys Cys
                560                 565                 570
Gly His Thr Gly Arg Cys Trp Ile Tyr Asn Lys Thr Lys Met Ala
                575                 580                 585
Phe Leu Leu Val Gly Ile Cys Phe Leu Cys Lys Leu Cys Thr Ile
                590                 595                 600
Ile Phe Thr Thr Ile Ala Phe Phe Ile Tyr Lys Arg Arg Leu Asn
                605                 610                 615
Glu Asn Thr Asp Phe Pro Asp Val Thr Val Lys Asn Pro Lys Val
                620                 625                 630
Lys Lys Lys Glu Glu Thr Asp Leu
                635
```

<210> 16
<211> 507
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 90011608CD1

<400> 16

```
Met Thr Ala Ser Thr Pro Glu Ala Thr Pro Asn Met Glu Leu Lys
  1               5                  10                  15
Ala Pro Ala Ala Gly Gly Leu Asn Ala Gly Pro Val Pro Pro Ala
                 20                  25                  30
Ala Leu Ser Thr Gln Arg Leu Arg Asn Glu Asp Tyr His Asp Tyr
                 35                  40                  45
Ser Ser Thr Asp Val Ser Pro Glu Glu Ser Pro Ser Glu Gly Leu
                 50                  55                  60
Asn Asn Leu Ser Ser Pro Gly Ser Tyr Gln Arg Phe Gly Gln Ser
                 65                  70                  75
Asn Ser Thr Thr Trp Phe Gln Thr Leu Ile His Leu Leu Lys Gly
                 80                  85                  90
```

```
Asn Ile Gly Thr Gly Leu Leu Gly Leu Pro Leu Ala Val Lys Asn
                 95                  100                 105
Ala Gly Ile Val Met Gly Pro Ile Ser Leu Leu Ile Ile Gly Ile
                110                 115                 120
Val Ala Val His Cys Met Gly Ile Leu Val Lys Cys Ala His His
                125                 130                 135
Phe Cys Arg Arg Leu Asn Lys Ser Phe Val Asp Tyr Gly Asp Thr
                140                 145                 150
Val Met Tyr Gly Leu Glu Ser Ser Pro Cys Ser Trp Leu Arg Asn
                155                 160                 165
His Ala His Trp Gly Arg Arg Val Val Asp Phe Phe Leu Ile Val
                170                 175                 180
Thr Gln Leu Gly Phe Cys Cys Val Tyr Phe Val Phe Leu Ala Asp
                185                 190                 195
Asn Phe Lys Gln Val Ile Glu Ala Ala Asn Gly Thr Thr Asn Asn
                200                 205                 210
Cys His Asn Asn Glu Thr Val Ile Leu Thr Pro Thr Met Asp Ser
                215                 220                 225
Arg Leu Tyr Met Leu Ser Phe Leu Pro Phe Leu Val Leu Leu Val
                230                 235                 240
Phe Ile Arg Asn Leu Arg Ala Leu Ser Ile Phe Ser Leu Leu Ala
                245                 250                 255
Asn Ile Thr Met Leu Val Ser Leu Val Met Ile Tyr Gln Phe Ile
                260                 265                 270
Val Gln Arg Ile Pro Asp Pro Ser His Leu Pro Leu Val Ala Pro
                275                 280                 285
Trp Lys Thr Tyr Pro Leu Phe Phe Gly Thr Ala Ile Phe Ser Phe
                290                 295                 300
Glu Gly Ile Gly Met Val Leu Pro Leu Glu Asn Lys Met Lys Asp
                305                 310                 315
Pro Arg Lys Phe Pro Leu Ile Leu Tyr Leu Gly Met Val Ile Val
                320                 325                 330
Thr Ile Leu Tyr Ile Ser Leu Gly Cys Leu Gly Tyr Leu Gln Phe
                335                 340                 345
Gly Ala Asn Ile Gln Gly Ser Ile Thr Leu Asn Leu Pro Asn Cys
                350                 355                 360
Trp Leu Tyr Gln Ser Val Lys Leu Leu Tyr Ser Ile Gly Ile Phe
                365                 370                 375
Phe Thr Tyr Ala Leu Gln Phe Tyr Val Pro Ala Glu Ile Ile Ile
                380                 385                 390
Pro Phe Phe Val Ser Arg Ala Pro Glu His Cys Glu Leu Val Val
                395                 400                 405
Asp Leu Phe Val Arg Thr Val Leu Val Cys Leu Thr Cys Ile Leu
                410                 415                 420
Ala Ile Leu Ile Pro Arg Leu Asp Leu Val Ile Ser Leu Val Gly
                425                 430                 435
Ser Val Ser Ser Ser Ala Leu Ala Leu Ile Ile Pro Pro Leu Leu
                440                 445                 450
Glu Val Thr Thr Phe Tyr Ser Glu Gly Met Ser Pro Leu Thr Ile
                455                 460                 465
Phe Lys Asp Ala Leu Ile Ser Ile Leu Gly Phe Val Gly Phe Val
                470                 475                 480
Val Gly Thr Tyr Glu Ala Leu Tyr Glu Leu Ile Gln Pro Ser Asn
                485                 490                 495
Ala Pro Ile Phe Ile Asn Ser Thr Cys Ala Phe Ile
                500                 505
```

<210> 17
<211> 568
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 90024583CD1

<400> 17

```
Met Ala Ser Ala Leu Ser Tyr Val Ser Lys Phe Lys Ser Phe Val
  1               5                  10                  15
Ile Leu Phe Val Thr Pro Leu Leu Leu Leu Pro Leu Val Ile Leu
             20                  25                  30
Met Pro Ala Lys Phe Val Arg Cys Ala Tyr Val Ile Ile Leu Met
         35                  40                  45
Ala Ile Tyr Trp Cys Thr Glu Val Ile Pro Leu Ala Val Thr Ser
     50                  55                  60
Leu Met Pro Val Leu Leu Phe Pro Leu Phe Gln Ile Leu Asp Ser
 65                  70                  75
Arg Gln Val Cys Val Gln Tyr Met Lys Asp Thr Asn Met Leu Phe
             80                  85                  90
Leu Gly Gly Leu Ile Val Ala Val Ala Val Glu Arg Trp Asn Leu
         95                 100                 105
His Lys Arg Ile Ala Leu Arg Thr Leu Leu Trp Val Gly Ala Lys
    110                 115                 120
Pro Ala Arg Leu Met Leu Gly Phe Met Gly Val Thr Ala Pro Leu
    125                 130                 135
Ser Met Trp Ile Ser Asn Thr Ala Thr Thr Ala Met Met Val Pro
    140                 145                 150
Ile Val Glu Ala Ile Leu Gln Gln Met Glu Ala Thr Ser Ala Ala
    155                 160                 165
Thr Glu Ala Gly Leu Glu Leu Val Asp Lys Gly Lys Ala Lys Glu
    170                 175                 180
Leu Pro Gly Ser Gln Val Ile Phe Glu Gly Pro Thr Leu Gly Gln
    185                 190                 195
Gln Glu Asp Gln Glu Arg Lys Arg Leu Cys Lys Ala Met Thr Leu
    200                 205                 210
Cys Ile Cys Tyr Ala Ala Ser Ile Gly Gly Thr Ala Thr Leu Thr
    215                 220                 225
Gly Thr Gly Pro Asn Val Val Leu Leu Gly Gln Met Asn Glu Leu
    230                 235                 240
Phe Pro Asp Ser Lys Asp Leu Val Asn Phe Ala Ser Trp Phe Ala
    245                 250                 255
Phe Ala Phe Pro Asn Met Leu Val Met Leu Leu Phe Ala Trp Leu
    260                 265                 270
Trp Leu Gln Phe Val Tyr Met Arg Phe Asn Phe Lys Lys Ser Trp
    275                 280                 285
Gly Cys Gly Leu Glu Ser Lys Lys Asn Glu Lys Ala Ala Leu Lys
    290                 295                 300
Val Leu Gln Glu Glu Tyr Arg Lys Leu Gly Pro Leu Ser Phe Ala
    305                 310                 315
Glu Ile Asn Val Leu Ile Cys Phe Phe Leu Leu Val Ile Leu Trp
    320                 325                 330
Phe Ser Arg Asp Pro Gly Phe Met Pro Gly Trp Leu Thr Val Ala
    335                 340                 345
```

```
Trp Val Glu Gly Glu Thr Lys Tyr Val Ser Asp Ala Thr Val Ala
                350                 355                 360
Ile Phe Val Ala Thr Leu Leu Phe Ile Val Pro Ser Gln Lys Pro
                365                 370                 375
Lys Phe Asn Phe Arg Ser Gln Thr Glu Glu Glu Arg Lys Thr Pro
                380                 385                 390
Phe Tyr Pro Pro Pro Leu Leu Asp Trp Lys Val Thr Gln Glu Lys
                395                 400                 405
Val Pro Trp Gly Ile Val Leu Leu Leu Gly Gly Gly Phe Ala Leu
                410                 415                 420
Ala Lys Gly Ser Glu Ala Ser Gly Leu Ser Val Trp Met Gly Lys
                425                 430                 435
Gln Met Glu Pro Leu His Ala Val Pro Pro Ala Ala Ile Thr Leu
                440                 445                 450
Ile Leu Ser Leu Leu Val Ala Val Phe Thr Glu Cys Thr Ser Asn
                455                 460                 465
Val Ala Thr Thr Thr Leu Phe Leu Pro Ile Phe Ala Ser Met Ser
                470                 475                 480
Arg Ser Asn Gly Leu Asn Pro Leu Tyr Ile Met Leu Pro Cys Thr
                485                 490                 495
Leu Ser Ala Ser Phe Ala Phe Met Leu Pro Val Ala Thr Pro Pro
                500                 505                 510
Asn Ala Ile Val Phe Thr Tyr Gly His Leu Lys Val Ala Asp Met
                515                 520                 525
Val Lys Thr Gly Val Ile Met Asn Ile Ile Gly Val Phe Cys Val
                530                 535                 540
Phe Leu Ala Val Asn Thr Trp Gly Arg Ala Ile Phe Asp Leu Asp
                545                 550                 555
His Phe Pro Asp Trp Ala Asn Val Thr His Ile Glu Thr
                560                 565
```

<210> 18
<211> 595
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 90113658CD1

<400> 18

```
Met Ile Leu Ile Lys Gln Lys Arg Leu Phe Pro Cys Trp Ile Pro
  1               5                  10                  15
Ala Leu Phe Ile Gly Phe Ser Gln Phe Ser Asp Ser Phe Leu Leu
                20                  25                  30
Asp Gln Pro Asn Phe Trp Cys Arg Gly Ala Gly Lys Gly Thr Glu
                35                  40                  45
Leu Ala Gly Val Thr Thr Thr Gly Arg Gly Gly Asp Met Gly Asn
                50                  55                  60
Trp Thr Ser Leu Pro Thr Thr Pro Phe Ala Thr Ala Pro Trp Glu
                65                  70                  75
Ala Ala Gly Asn Arg Ser Asn Ser Ser Gly Ala Asp Gly Gly Asp
                80                  85                  90
Thr Pro Pro Leu Pro Ser Pro Pro Asp Lys Gly Asp Asn Ala Ser
                95                 100                 105
Asn Cys Asp Cys Arg Ala Trp Asp Tyr Gly Ile Arg Ala Gly Leu
```

```
                110                 115                 120
Val Gln Asn Val Val Ser Lys Trp Asp Leu Val Cys Asp Asn Ala
                125                 130                 135
Trp Lys Val His Ile Ala Lys Phe Ser Leu Leu Val Gly Leu Ile
                140                 145                 150
Phe Gly Tyr Leu Ile Thr Gly Cys Ile Ala Asp Trp Val Gly Arg
                155                 160                 165
Arg Pro Val Leu Leu Phe Ser Ile Ile Phe Ile Leu Ile Phe Gly
                170                 175                 180
Leu Thr Val Ala Leu Ser Val Asn Val Thr Met Phe Ser Thr Leu
                185                 190                 195
Arg Phe Phe Glu Gly Phe Cys Leu Ala Gly Ile Ile Leu Thr Leu
                200                 205                 210
Tyr Ala Leu Arg Ile Glu Leu Cys Pro Pro Gly Lys Arg Phe Met
                215                 220                 225
Ile Thr Met Val Ala Ser Phe Val Ala Met Ala Gly Gln Phe Leu
                230                 235                 240
Met Pro Gly Leu Ala Ala Leu Cys Arg Asp Trp Gln Val Leu Gln
                245                 250                 255
Ala Leu Ile Ile Cys Pro Phe Leu Leu Met Leu Leu Tyr Trp Ser
                260                 265                 270
Ile Phe Pro Glu Ser Leu Arg Trp Leu Met Ala Thr Gln Gln Phe
                275                 280                 285
Glu Ser Ala Lys Arg Leu Ile Leu His Phe Thr Gln Lys Asn Arg
                290                 295                 300
Met Asn Pro Glu Gly Asp Ile Lys Gly Val Ile Pro Glu Leu Glu
                305                 310                 315
Lys Glu Leu Ser Arg Arg Pro Lys Lys Val Cys Ile Val Lys Val
                320                 325                 330
Val Gly Thr Arg Asn Leu Trp Lys Asn Ile Val Val Leu Cys Val
                335                 340                 345
Asn Ser Leu Thr Gly Tyr Gly Ile His His Cys Phe Ala Arg Ser
                350                 355                 360
Met Met Gly His Glu Val Lys Val Pro Leu Leu Glu Asn Phe Tyr
                365                 370                 375
Ala Asp Tyr Tyr Thr Thr Ala Ser Ile Ala Leu Val Ser Cys Leu
                380                 385                 390
Ala Met Cys Val Val Val Arg Phe Leu Gly Arg Arg Gly Gly Leu
                395                 400                 405
Leu Leu Phe Met Ile Leu Thr Ala Leu Ala Ser Leu Leu Gln Leu
                410                 415                 420
Gly Leu Leu Asn Leu Ile Gly Lys Tyr Ser Gln His Pro Asp Ser
                425                 430                 435
Gly Met Ser Asp Ser Val Lys Asp Lys Phe Ser Ile Ala Phe Ser
                440                 445                 450
Ile Val Gly Met Phe Ala Ser His Ala Val Gly Ser Leu Ser Val
                455                 460                 465
Phe Phe Cys Ala Glu Ile Thr Pro Thr Val Ile Arg Cys Gly Gly
                470                 475                 480
Leu Gly Leu Val Leu Ala Ser Ala Gly Phe Gly Met Leu Thr Ala
                485                 490                 495
Pro Ile Ile Glu Leu His Asn Gln Lys Gly Tyr Phe Leu His His
                500                 505                 510
Ile Ile Phe Ala Cys Cys Thr Leu Ile Cys Ile Ile Cys Ile Leu
                515                 520                 525
Leu Leu Pro Glu Ser Arg Asp Gln Asn Leu Pro Glu Asn Ile Ser
```

```
                530                 535                 540
Asn Gly Glu His Tyr Thr Arg Gln Pro Leu Leu Pro His Lys Lys
                545                 550                 555
Gly Glu Gln Pro Leu Leu Leu Thr Asn Ala Glu Leu Lys Asp Tyr
                560                 565                 570
Ser Gly Leu His Asp Ala Ala Ala Ala Gly Asp Thr Leu Pro Glu
                575                 580                 585
Gly Ala Thr Ala Asn Gly Met Lys Ala Met
                590                 595

<210> 19
<211> 602
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3942766CD1

<400> 19
Met Ala Ala Leu Ala Ala Ala Ala Lys Lys Val Trp Ser Ala Arg
  1               5                  10                  15
Arg Leu Leu Val Leu Leu Phe Thr Pro Leu Ala Leu Leu Pro Val
                20                  25                  30
Val Phe Ala Leu Pro Pro Lys Glu Gly Arg Cys Leu Phe Val Ile
                35                  40                  45
Leu Leu Met Ala Val Tyr Trp Cys Thr Glu Ala Leu Pro Leu Ser
                50                  55                  60
Val Thr Ala Leu Leu Pro Ile Val Leu Phe Pro Phe Met Gly Ile
                65                  70                  75
Leu Pro Ser Asn Lys Val Cys Pro Gln Tyr Phe Leu Asp Thr Asn
                80                  85                  90
Phe Leu Phe Leu Ser Gly Leu Ile Met Ala Ser Ala Ile Glu Glu
                95                 100                 105
Trp Asn Leu His Arg Arg Ile Ala Leu Lys Ile Leu Met Leu Val
               110                 115                 120
Gly Val Gln Pro Ala Arg Leu Ile Leu Gly Met Met Val Thr Thr
               125                 130                 135
Ser Phe Leu Ser Met Trp Leu Ser Asn Thr Ala Ser Thr Ala Met
               140                 145                 150
Met Leu Pro Ile Ala Asn Ala Ile Leu Lys Ser Leu Phe Gly Gln
               155                 160                 165
Lys Glu Val Arg Lys Asp Pro Ser Gln Glu Ser Glu Glu Asn Thr
               170                 175                 180
Ala Ala Val Arg Arg Asn Gly Leu His Thr Val Pro Thr Glu Met
               185                 190                 195
Gln Phe Leu Ala Ser Thr Glu Ala Lys Asp His Pro Gly Glu Thr
               200                 205                 210
Glu Val Pro Leu Asp Leu Pro Ala Asp Ser Arg Lys Glu Asp Glu
               215                 220                 225
Tyr Arg Arg Asn Ile Trp Lys Gly Phe Leu Ile Ser Ile Pro Tyr
               230                 235                 240
Ser Ala Ser Ile Gly Gly Thr Ala Thr Leu Thr Gly Thr Ala Pro
               245                 250                 255
Asn Leu Ile Leu Leu Gly Gln Leu Lys Ser Phe Phe Pro Gln Cys
               260                 265                 270
```

```
Asp Val Val Asn Phe Gly Ser Trp Phe Ile Phe Ala Phe Pro Leu
                275                 280                 285
Met Leu Leu Phe Leu Leu Ala Gly Trp Leu Trp Ile Ser Phe Leu
                290                 295                 300
Tyr Gly Gly Leu Ser Phe Arg Gly Trp Arg Lys Asn Lys Ser Glu
                305                 310                 315
Ile Arg Thr Asn Ala Glu Asp Arg Ala Arg Ala Val Ile Arg Glu
                320                 325                 330
Glu Tyr Gln Asn Leu Gly Pro Ile Lys Phe Ala Glu Gln Ala Val
                335                 340                 345
Phe Ile Leu Phe Cys Met Phe Ala Ile Leu Leu Phe Thr Arg Asp
                350                 355                 360
Pro Lys Phe Ile Pro Gly Trp Ala Ser Leu Phe Asn Pro Gly Phe
                365                 370                 375
Leu Ser Asp Ala Val Thr Gly Val Ala Ile Val Thr Ile Leu Phe
                380                 385                 390
Phe Phe Pro Ser Gln Arg Pro Ser Leu Lys Trp Trp Phe Asp Phe
                395                 400                 405
Lys Ala Pro Asn Thr Glu Thr Glu Pro Leu Leu Thr Trp Lys Lys
                410                 415                 420
Ala Gln Glu Thr Val Pro Trp Asn Ile Ile Leu Leu Leu Gly Gly
                425                 430                 435
Gly Phe Ala Met Ala Lys Gly Cys Glu Glu Ser Gly Leu Ser Val
                440                 445                 450
Trp Ile Gly Gly Gln Leu His Pro Leu Glu Asn Val Pro Pro Ala
                455                 460                 465
Leu Ala Val Leu Leu Ile Thr Val Val Ile Ala Phe Phe Thr Glu
                470                 475                 480
Phe Ala Ser Asn Thr Ala Thr Ile Ile Ile Phe Leu Pro Val Leu
                485                 490                 495
Ala Glu Leu Ala Ile Arg Leu Arg Val His Pro Leu Tyr Leu Met
                500                 505                 510
Ile Pro Gly Thr Val Gly Cys Ser Phe Ala Phe Met Leu Pro Val
                515                 520                 525
Ser Thr Pro Pro Asn Ser Ile Ala Phe Ala Ser Gly His Leu Leu
                530                 535                 540
Val Lys Asp Met Val Arg Thr Gly Leu Leu Met Asn Leu Met Gly
                545                 550                 555
Val Leu Leu Leu Ser Leu Ala Met Asn Thr Trp Ala Gln Thr Ile
                560                 565                 570
Phe Gln Leu Gly Thr Phe Pro Asp Trp Ala Asp Met Tyr Ser Val
                575                 580                 585
Asn Val Thr Ala Leu Pro Pro Thr Leu Ala Asn Asp Thr Phe Arg
                590                 595                 600
Thr Leu
```

<210> 20
<211> 372
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7501987CD1

<400> 20

```
Met Val Pro Ala Gly Trp Val Arg Gly Leu Glu Leu Ser Leu Trp
  1               5                  10                  15
Gly Gly Asp Pro Val Val Pro Trp Ser Cys Arg Phe Cys Ser Gln
             20                  25                  30
Gln Asp Asp Gly Gln Asp Arg Glu Arg Leu Thr Tyr Phe Gln Asn
             35                  40                  45
Leu Pro Glu Ser Leu Thr Ser Leu Leu Val Leu Leu Thr Thr Ala
             50                  55                  60
Asn Asn Pro Asp Val Met Ile Pro Ala Tyr Ser Lys Asn Arg Ala
             65                  70                  75
Tyr Ala Ile Phe Phe Ile Val Phe Thr Val Ile Gly Ser Leu Phe
             80                  85                  90
Leu Met Asn Leu Leu Thr Ala Ile Ile Tyr Ser Gln Phe Arg Gly
             95                 100                 105
Tyr Leu Met Lys Ser Leu Gln Thr Ser Leu Phe Arg Arg Arg Leu
            110                 115                 120
Gly Thr Arg Ala Ala Phe Glu Val Leu Ser Ser Met Val Gly Glu
            125                 130                 135
Gly Gly Ala Phe Pro Gln Ala Val Gly Val Lys Pro Gln Asn Leu
            140                 145                 150
Leu Gln Val Leu Gln Lys Val Gln Leu Asp Ser Ser His Lys Gln
            155                 160                 165
Ala Met Met Glu Lys Val Arg Ser Tyr Gly Ser Val Leu Leu Ser
            170                 175                 180
Ala Glu Glu Phe Gln Lys Leu Phe Asn Glu Leu Asp Arg Ser Val
            185                 190                 195
Val Lys Glu His Pro Pro Arg Pro Glu Tyr Gln Ser Pro Phe Leu
            200                 205                 210
Gln Ser Ala Gln Phe Leu Phe Gly His Tyr Tyr Phe Asp Tyr Leu
            215                 220                 225
Gly Asn Leu Ile Ala Leu Ala Asn Leu Val Ser Ile Cys Val Phe
            230                 235                 240
Leu Val Leu Asp Ala Asp Val Leu Pro Ala Glu Arg Asp Asp Phe
            245                 250                 255
Ile Leu Gly Ile Leu Asn Cys Val Phe Ile Val Tyr Tyr Leu Leu
            260                 265                 270
Glu Leu Leu Leu Lys Val Phe Ala Leu Gly Leu Arg Gly Tyr Leu
            275                 280                 285
Ser Tyr Pro Ser Asn Val Phe Asp Gly Leu Leu Thr Val Val Leu
            290                 295                 300
Leu Glu Ala Gly Asp Gly Gly Pro Ala Val Ala Val Gly His Asp
            305                 310                 315
Pro His Ala Glu His Ala His Arg Val Pro Leu Pro Ala Tyr His
            320                 325                 330
Pro Gln His Glu Ala Asp Gly Arg Gly Gly Gln Tyr Arg Pro Gly
            335                 340                 345
Pro Gly Ala Glu His Ala Cys Val Trp Arg Asp Pro Gly Gly Gly
            350                 355                 360
Leu Leu Arg Ile Cys His His Trp Asp Gln Leu Val
            365                 370
```

<210> 21
<211> 165
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503223CD1

<400> 21
Met Thr Leu Leu Pro Gly Asp Asn Ser Asp Tyr Asp Tyr Ser Ala
1 5 10 15
Leu Ser Cys Thr Ser Asp Ala Ser Phe His Pro Ala Phe Leu Pro
20 25 30
Gln Arg Gln Ala Ile Lys Gly Ala Phe Tyr Arg Arg Ala Gln Arg
35 40 45
Leu Arg Pro Gln Asp Glu Pro Arg Gln Gly Cys Gln Pro Glu Asp
50 55 60
Arg Arg Arg Arg Ile Ile Ile Asn Val Gly Gly Ile Lys Tyr Ser
65 70 75
Leu Pro Trp Thr Thr Leu Asp Glu Phe Pro Leu Thr Arg Leu Gly
80 85 90
Gln Leu Lys Ala Cys Thr Asn Phe Asp Asp Ile Leu Asn Val Cys
95 100 105
Asp Asp Tyr Asp Val Thr Cys Asn Glu Phe Phe Phe Asp Arg Asn
110 115 120
Pro Gly Ala Phe Gly Thr Ile Leu Thr Phe Leu Arg Ala Gly Lys
125 130 135
Leu Arg Leu Leu Arg Glu Met Cys Ala Leu Ser Phe Gln Asp Ser
140 145 150
Asp Ile Leu Phe Gly Ser Ala Ser Ser Asp Thr Arg Asp Asn Asn
155 160 165

<210> 22
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503566CD1

<400> 22
Met Leu Arg Thr Ile Leu Asp Ala Pro Gln Arg Leu Leu Lys Glu
1 5 10 15
Gly Arg Ala Ser Arg Gln Leu Val Leu Val Val Val Phe Val Ala
20 25 30
Leu Leu Leu Asp Asn Met Leu Phe Thr Val Val Val Pro Ile Val
35 40 45
Pro Thr Phe Leu Tyr Asp Met Glu Phe Lys Glu Val Asn Ser Ser
50 55 60
Leu His Leu Gly His Ala Gly Ser Ser Pro His Ala Leu Ala Ser
65 70 75
Pro Ala Phe Ser Thr Ile Phe Ser Phe Phe Asn Asn Asn Thr Val
80 85 90
Ala Val Glu Glu Ser Val Pro Ser Gly Ile Ala Trp Met Asn Asp
95 100 105
Thr Ala Ser Thr Ile Pro Pro Pro Ala Thr Glu Ala Ile Ser Ala
110 115 120
His Lys Asn Asn Cys Leu Gln Gly Thr Gly Phe Leu Glu Glu Glu

```
                125                 130                 135
Thr Thr Arg Val Gly Val Leu Phe Ala Ser Lys Ala Val Met Gln
                140                 145                 150
Leu Leu Val Asn Pro Phe Val Gly Pro Leu Thr Asn Arg Ile Gly
                155                 160                 165
Tyr His Ile Pro Met Phe Ala Gly Phe Val Ile Met Phe Leu Ser
                170                 175                 180
Thr Val Ser Leu Gly Met Leu Ala Ser Val Tyr Thr Asp Asp His
                185                 190                 195
Glu Arg Gly Arg Ala Met Gly Thr Ala Leu Gly Gly Leu Ala Leu
                200                 205                 210
Gly Leu Leu Val Gly Ala Pro Phe Gly Ser Val Met Tyr Glu Phe
                215                 220                 225
Val Gly Lys Ser Ala Pro Phe Leu Ile Leu Ala Phe Leu Ala Leu
                230                 235                 240
Leu Asp Gly Ala Leu Gln Leu Cys Ile Leu Gln Pro Ser Lys Val
                245                 250                 255
Ser Pro Glu Ser Ala Lys Gly Thr Pro Leu Phe Met Leu Leu Lys
                260                 265                 270
Asp Pro Tyr Ile Leu Val Ala Ala Gly Ser Ile Cys Phe Ala Asn
                275                 280                 285
Met Gly Val Ala Ile Leu Glu Pro Thr Leu Pro Ile Trp Met Met
                290                 295                 300
Gln Thr Met Cys Ser Pro Lys Trp Gln Leu Gly Leu Ala Phe Leu
                305                 310                 315
Pro Ala Ser Val Ser Tyr Leu Ile Gly Thr Asn Leu Phe Gly Val
                320                 325                 330
Leu Ala Asn Lys Met Gly Arg Trp Leu Cys Ser Leu Ile Gly Met
                335                 340                 345
Leu Val Val Gly Thr Ser Leu Leu Cys Val Pro Leu Ala His Asn
                350                 355                 360
Ile Phe Gly Leu Ile Gly Pro Asn Ala Gly Leu Gly Leu Ala Ile
                365                 370                 375
Gly Met Val Asp Ser Ser Met Met Pro Ile Met Gly His Leu Val
                380                 385                 390
Asp Leu Arg His Thr Ser Val Tyr Gly Ser Val Tyr Ala Ile Ala
                395                 400                 405
Asp Val Ala Phe Cys Met Gly Phe Ala Ile Gly Pro Ser Thr Gly
                410                 415                 420
Gly Ala Ile Val Lys Ala Ile Gly Phe Pro Trp Leu Met Val Ile
                425                 430                 435
Thr Gly Val Ile Asn Ile Val Tyr Ala Pro Leu Cys Tyr Tyr Leu
                440                 445                 450
Arg Ser Pro Pro Ala Lys Glu Glu Lys Leu Ala Ile Leu Ser Gln
                455                 460                 465
Asp Cys Pro Met Glu Thr Arg Met Tyr Ala Thr Gln Lys Pro Thr
                470                 475                 480
Lys Glu Phe Pro Leu Gly Glu Asp Ser Asp Glu Glu Pro Asp His
                485                 490                 495
Glu Glu
```

<210> 23
<211> 67
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7505122CD1

<400> 23
Met Gln Lys Val Thr Leu Gly Leu Leu Val Phe Leu Ala Gly Phe
1 5 10 15
Pro Val Leu Asp Ala Asn Asp Leu Glu Asp Lys Asn Ser Pro Phe
20 25 30
Tyr Tyr Asp Trp His Ser Leu Gln Val Gly Gly Leu Ile Cys Ala
35 40 45
Gly Val Leu Cys Met Ala Gly Pro His Leu Thr Ser Arg Lys Arg
50 55 60
Val Ser Leu Phe Asn Phe Phe
65

<210> 24
<211> 152
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511620CD1

<400> 24
Met Gly Leu Ala Asp Ala Ser Gly Pro Arg Asp Thr Gln Ala Leu
1 5 10 15
Leu Ser Ala Thr Gln Ala Met Asp Leu Arg Arg Arg Asp Tyr His
20 25 30
Met Glu Arg Pro Leu Leu Asn Gln Glu His Leu Glu Glu Leu Gly
35 40 45
Arg Trp Gly Ser Ala Pro Arg Thr His Gln Trp Arg Thr Trp Leu
50 55 60
Gln Cys Ser Arg Ala Arg Ala Tyr Ala Leu Leu Leu Gln His Leu
65 70 75
Pro Val Leu Val Trp Leu Pro Arg Tyr Pro Val Arg Asp Trp Leu
80 85 90
Leu Gly Asp Leu Leu Ser Gly Leu Ser Val Ala Ile Met Gln Leu
95 100 105
Pro Gln Gly Leu Ala Tyr Ala Leu Leu Ala Gly Leu Pro Pro Val
110 115 120
Phe Gly Leu Tyr Ser Ser Phe Tyr Pro Val Phe Ile Tyr Phe Leu
125 130 135
Phe Gly Thr Ser Arg His Ile Ser Val Gly Leu Glu Arg Leu His
140 145 150
Asp Gln

<210> 25
<211> 467
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7506995CD1

<400> 25

```
Met Val Pro Ala Gly Trp Val Arg Gly Leu Glu Leu Ser Leu Trp
  1               5                  10                  15
Gly Gly Asp Pro Val Val Pro Trp Ser Cys Arg Phe Cys Ser Gln
             20                  25                  30
Gln Asp Asp Gly Gln Asp Arg Glu Arg Leu Thr Tyr Phe Gln Asn
             35                  40                  45
Leu Pro Glu Ser Leu Thr Ser Leu Leu Val Leu Leu Thr Thr Ala
             50                  55                  60
Asn Asn Pro Asp Val Met Ile Pro Ala Tyr Ser Lys Asn Arg Ala
             65                  70                  75
Tyr Ala Ile Phe Phe Ile Val Phe Thr Val Ile Gly Ser Leu Phe
             80                  85                  90
Leu Met Asn Leu Leu Thr Ala Ile Ile Tyr Ser Gln Phe Arg Gly
             95                 100                 105
Tyr Leu Met Lys Ser Leu Gln Thr Ser Leu Phe Arg Arg Arg Leu
            110                 115                 120
Gly Thr Arg Ala Ala Phe Glu Val Leu Ser Ser Met Val Gly Glu
            125                 130                 135
Gly Gly Ala Phe Pro Gln Ala Thr Arg Arg Gly Pro Ser Thr Ser
            140                 145                 150
Leu Arg Phe Cys Arg Ala Pro Ser Ser Ser Ser Ala Thr Thr Thr
            155                 160                 165
Leu Thr Thr Trp Gly Thr Ser Ser Pro Trp Gln Thr Trp Cys Pro
            170                 175                 180
Phe Ala Cys Ser Trp Cys Trp Met Gln Met Cys Cys Leu Leu Ser
            185                 190                 195
Val Met Thr Ser Ser Trp Gly Phe Ser Thr Ala Ser Ser Leu Cys
            200                 205                 210
Thr Thr Cys Trp Arg Cys Cys Ser Arg Ser Leu Pro Trp Ala Cys
            215                 220                 225
Glu Gly Thr Cys Pro Thr Pro Ala Thr Cys Leu Thr Gly Ser Ser
            230                 235                 240
Pro Leu Ser Cys Trp Arg Pro Glu Met Val Gly Leu Leu Ser Leu
            245                 250                 255
Trp Asp Met Thr Arg Met Leu Asn Met Leu Ile Val Phe Arg Phe
            260                 265                 270
Leu Arg Ile Ile Pro Ser Met Lys Pro Met Ala Val Val Ala Ser
            275                 280                 285
Thr Val Leu Gly Leu Val Gln Asn Met Arg Ala Phe Gly Gly Ile
            290                 295                 300
Leu Val Val Val Tyr Tyr Val Phe Ala Ile Ile Gly Ile Asn Leu
            305                 310                 315
Phe Arg Gly Val Ile Val Ala Leu Pro Gly Asn Ser Ser Leu Ala
            320                 325                 330
Pro Ala Asn Gly Ser Ala Pro Cys Gly Ser Phe Glu Gln Leu Glu
            335                 340                 345
Tyr Trp Ala Asn Asn Phe Asp Asp Phe Ala Ala Ala Leu Val Thr
            350                 355                 360
Leu Trp Asn Leu Met Val Val Asn Asn Trp Gln Val Phe Leu Asp
            365                 370                 375
Ala Tyr Arg Arg Tyr Ser Gly Pro Trp Ser Lys Ile Tyr Phe Val
            380                 385                 390
Leu Trp Trp Leu Val Ser Ser Val Ile Trp Val Asn Leu Phe Leu
```

```
                395                 400                 405
Ala Leu Ile Leu Glu Asn Phe Leu His Lys Trp Asp Pro Arg Ser
                410                 415                 420
His Leu Gln Pro Leu Ala Gly Thr Pro Glu Ala Thr Tyr Gln Met
                425                 430                 435
Thr Val Glu Leu Leu Phe Arg Asp Ile Leu Glu Glu Pro Gly Glu
                440                 445                 450
Asp Glu Leu Thr Glu Arg Leu Ser Gln His Pro His Leu Trp Leu
                455                 460                 465
Cys Arg
```

<210> 26
<211> 490
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506996CD1

<400> 26

```
Met Val Pro Ala Gly Trp Val Arg Gly Leu Glu Leu Ser Leu Trp
  1                 5                  10                  15
Gly Gly Asp Pro Val Val Pro Trp Ser Cys Arg Phe Cys Ser Gln
                 20                  25                  30
Gln Asp Asp Gly Gln Asp Arg Glu Arg Leu Thr Tyr Phe Gln Asn
                 35                  40                  45
Leu Pro Glu Ser Leu Thr Ser Leu Leu Val Leu Leu Thr Thr Ala
                 50                  55                  60
Asn Asn Pro Asp Val Met Ile Pro Ala Tyr Ser Lys Asn Arg Ala
                 65                  70                  75
Tyr Ala Ile Phe Phe Ile Val Phe Thr Val Ile Gly Ser Leu Phe
                 80                  85                  90
Leu Met Asn Leu Leu Thr Ala Ile Ile Tyr Ser Gln Phe Arg Gly
                 95                 100                 105
Tyr Leu Met Lys Ser Leu Gln Thr Ser Leu Phe Arg Arg Arg Leu
                110                 115                 120
Gly Thr Arg Ala Ala Phe Glu Val Leu Ser Ser Met Val Gly Glu
                125                 130                 135
Gly Gly Ala Phe Pro Gln Ala Val Gly Val Lys Pro Gln Asn Leu
                140                 145                 150
Leu Gln Val Leu Gln Lys Val Gln Leu Asp Ser Ser His Lys Gln
                155                 160                 165
Ala Met Met Glu Lys Val Arg Ser Tyr Gly Ser Val Leu Leu Ser
                170                 175                 180
Ala Glu Glu Phe Gln Lys Leu Phe Asn Glu Leu Asp Arg Ser Val
                185                 190                 195
Val Lys Glu His Pro Pro Arg Pro Glu Tyr Gln Ser Pro Phe Leu
                200                 205                 210
Gln Ser Ala Gln Phe Leu Phe Gly His Tyr Tyr Phe Asp Tyr Leu
                215                 220                 225
Gly Asn Leu Ile Ala Leu Ala Asn Leu Val Ser Ile Cys Val Phe
                230                 235                 240
Leu Val Leu Asp Ala Asp Val Leu Pro Ala Glu Arg Asp Asp Phe
                245                 250                 255
```

```
Ile Leu Gly Ile Leu Asn Cys Val Phe Ile Val Tyr Tyr Leu Leu
                260                 265                 270
Glu Leu Leu Leu Lys Val Phe Ala Leu Gly Leu Arg Gly Tyr Leu
                275                 280                 285
Ser Tyr Pro Ser Asn Val Phe Asp Gly Leu Leu Thr Val Val Leu
                290                 295                 300
Leu Pro Met Ala Val Val Ala Ser Thr Val Leu Gly Leu Val Gln
                305                 310                 315
Asn Met Arg Ala Phe Gly Gly Ile Leu Val Val Val Tyr Tyr Val
                320                 325                 330
Phe Ala Ile Ile Gly Ile Asn Leu Phe Arg Gly Val Ile Val Ala
                335                 340                 345
Leu Pro Gly Asn Ser Ser Leu Ala Pro Ala Asn Gly Ser Ala Pro
                350                 355                 360
Cys Gly Ser Phe Glu Gln Leu Glu Tyr Trp Ala Asn Asn Phe Asp
                365                 370                 375
Asp Phe Ala Ala Ala Leu Val Thr Leu Trp Asn Leu Met Val Val
                380                 385                 390
Asn Asn Trp Gln Val Phe Leu Asp Ala Tyr Arg Arg Tyr Ser Gly
                395                 400                 405
Pro Trp Ser Lys Ile Tyr Phe Val Leu Trp Trp Leu Val Ser Ser
                410                 415                 420
Val Ile Trp Val Asn Leu Phe Leu Ala Leu Ile Leu Glu Asn Phe
                425                 430                 435
Leu His Lys Trp Asp Pro Arg Ser His Leu Gln Pro Leu Ala Gly
                440                 445                 450
Thr Pro Glu Ala Thr Tyr Gln Met Thr Val Glu Leu Leu Phe Arg
                455                 460                 465
Asp Ile Leu Glu Glu Pro Gly Glu Asp Glu Leu Thr Glu Arg Leu
                470                 475                 480
Ser Gln His Pro His Leu Trp Leu Cys Arg
                485                 490
```

<210> 27
<211> 2343
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1853191CB1

<400> 27

```
cgcgtacagg aggcggcggc ggctcccagt caccggcccc cgccggcgag cgcacgatgc 60
actgcctggg cgccgagtac ctggtttctg cagaaggagc ccctaggcaa agggagtggc 120
gaccccagat ttataggaaa tgcacagata cggcatggtt attcctgttc tttctctttt 180
ggactggttt ggtgtttatc atgggctact cggtggtggc tggagccgcg ggaagactcc 240
tctttggcta tgacagcttt ggcaacatgt gtggcaagaa gaactccccc gtggaagggg 300
cccctctttc agggcaggac atgaccctaa aaaaacacgt gttctttatg aattcctgca 360
acctggaagt caaaggtacg cagctcaacc gcatggccct ctgtgtatcc aactgccctg 420
aagagcagct tgactccctg gaagaggtcc agttctttgc aaacaccagt gggtccttcc 480
tgtgtgttta tagtttgaat tccttcaact atacccacag tccaaaagca gactcactgt 540
gtcccaggct accagttcct ccaagcaagt catttccctt atttaaccga tgtgtccctc 600
aaacacctga gtgctactcc ctatttgcat ctgttttgat aaatgatgtt gacaccctcc 660
accgaattct aagtggaatc atgtcgggaa gagatacaat ccttggcctg tgtatcctcg 720
cattagcctt gtctttggcc atgatgttta ccttcagatt catcaccacc cttctggttc 780
```

acattttcat ttcattggtt attttgggat tgttgtttgt ctgcggtgtt ttatggtggc 840
tgtattatga ctataccaac gacctcagca tagaattgga cacagaaagg gaaaatatga 900
agtgcgtgct ggggtttgct atcgtatcca caggcatcac ggcagtgctg ctcgtcttga 960
tttttgttct cagaaagaga ataaaattga cagttgagct tttccaaatc acaaataaag 1020
ccatcagcag tgctcccttc ctgctgttcc agccactgtg gacatttgcc atcctcattt 1080
tcttctgggt cctctgggtg gctgtgctgc tgagcctggg aactgcagga gctgcccagg 1140
ttatggaagg cggccaagtg gaatataagc ccctttcggg cattcggtac atgtggtcgt 1200
accatttaat tggcctcatc tggactagtg aattcatcct tgcgtgccag caaatgacta 1260
tagctggggc agtggttact tgttatttca acagaagtaa aaatgatcct cctgatcatc 1320
ccatcctttc gtctctctcc attctcttct tctaccatca aggaaccatt gtgaaagggt 1380
catttttaat ctctgtggtg aggattccga gaatcattgt catgtacatg caaaacgcac 1440
tgaaagaaca gcatggtgca ttgtccaggt acctgttccg atgctgctac tgctgtttct 1500
ggtgtcttga caaatacctg ctccatctca accagaatgc atatactaca actgctatta 1560
atgggacaga tttctgtaca tcagcaaaag atgcattcaa aatcttgtcc aagaactcaa 1620
gtcactttac atctattaac tgctttggag acttcataat ttttctagga aaggtgttag 1680
tggtgtgttt cactgttttt ggaggactca tggcttttaa ctacaatcgg gcattccagg 1740
tgtgggcagt ccctctgtta ttggtagctt tttttgccta cttagtagcc catagttttt 1800
tatctgtgtt tgaaactgtg ctggatgcac ttttcctgtg ttttgctgtt gatctggaaa 1860
caaatgatgg atcgtcagaa aagccctact ttatggatca agaatttctg agtttcgtaa 1920
aaaggagcaa caaattaaac aatgcaaggg cacagcagga caagcactca ttaaggaatg 1980
aggagggaac agaactccag gccattgtga gatagatacc catttaggta tctgtacctg 2040
gaaaacattt ccttctaaga gccatttaca gaatagaaga tgagaccact agagaaaagt 2100
tagtgaattt ttttttaaaa gacctaataa accctattct tcctcattgt ctttgtcatt 2160
attgtttgac caggtaacaa tactggaact atattagttt acctttttt gtacaaatta 2220
ggacagaaaa actcttctaa aaccatgttt atatgcatca acttacaaag tacactatgt 2280
aagaactgag gtaagtttgt aagtgcacaa ctaataaata aacctttta agataaggaa 2340
aaa 2343

<210> 28
<211> 3145
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7497369CB1

<400> 28
ctggtgtccc cggggtccct cggatagagg tggaaacagg tcggacaaga gctggccgtg 60
ggggaagagc aatggtgtgc ggaagtgggg agccctatct agggacaagc agaaaactta 120
gggtggcaga tccccccgac cctgtgcagg ggctgccgga gaccagagag gctgctgtgt 180
gacctggacc aagccccttg ccctctcatg cctcagtttc cccttctctc cacggaggga 240
ttgagcagga catggagggc cccttccagt agccattctg ggactccctg cacctgtccc 300
caggggaggg tcctgagctg gcctccacgt ggctgctgtt ccccatagtg gaaccttgtg 360
tgtggagacg gctggaaggt cccgctggag caggtgagcc acctcctggg ctggctgctg 420
ggctgtgtca tcctgggagc aggctgtgac cggtgaggcc ccttcccttt tcctgctgcc 480
cagcagccct cccctccgcc cactctcagg gttcagctag acctggtgct gccaggatga 540
aggttggacc ctgggttatc ctgaccctgt ccaagcatgt cccccgctgc ccaccaggtt 600
tggacgccgg gcagtttttg tggcctccct ggtgctgacc acaggcctgg gggccagtga 660
ggccctggct gccagcttcc ctaccctgct ggtcctgcgc ctactccacg ggggcacatt 720
ggcaggggcc ctcctcgccc tgtatctggc tcgtgagtac cctgggggct gctgtcactg 780
ggggaggggc agtgggtggc ccgcaggcct ctgaggctcc cttgccgagg gccccgagct 840
gcagggacag tgagcagtga gtcccttggg catcccgctc ctgggcaggt caccaatagg 900
tccccgcagt tcccaatgga actgttccag tcctcccccga ggcctccact tcaacctgtc 960
tgtgtctgcc caggcctgga gttgtgtgac cctccccacc gcctggcctt ctccatgggg 1020
gctggccttt tctcggtggt gggcaccctg ctgctgcccg gcctggctgc gcttgtgcag 1080

gactggcgtc ttctgcaggg gctgggcgcc ctgatgagtg gactcttgct gctcttttgg 1140
gggtaagtgt ggtgggagct gggccagcag ccctcatcgg ggctgactat gtagcttccc 1200
tctgcaccga cgtatcccat gccaggtaaa cgcatgggtt acagggtcac agagaactac 1260
agtgatgtcc tgcccctgag cctgggggtg gggtgggggc cctggccttt gcctctcctt 1320
ccaggaggag gtggagggag ccgtgggcat cctcaccaac gctgcaggtt cccggccctg 1380
ttccccgagt ctccctgctg gctgctggcc acaggtcagg tagctcgagc caggaagatc 1440
ctgtggcgct ttgcagaagc cagtggcgtg ggccccgggg acagtcccct ggaggagaac 1500
tccctggcta cagagctgac catgctgtct gcacggagcc cccagccccg gtaccactcc 1560
ccactggggc ttctgcgtac ccgagtcacc tggagaaacg ggcttatctt gggcttcagc 1620
tcgctggttg gtggaggcat cagagctagc ttccgccgca gcctggcacc tcaggtgccg 1680
accttctacc tgccctactt cctggaggcc ggcctggagg cggcagcctt ggtcttcctg 1740
ctcctgacgg cagattgctg tggacgccgc cccgtgctgc tgctgggcac catggtcaca 1800
ggcctggcat ccctgctgct cctcgctggg gcccagtatc tgccaggctg gactgtgctg 1860
ttcctctctg tcctggggct cctggcctcc cgggctgtgt ccgcactcag cagcctcttc 1920
gcggccgagg tcttccccac ggtgatcagg ggggccgggc tgggcctggt gctgggggcc 1980
gggttcctgg gccaggcagc cggccccctg gacaccctgc acggccggca gggcttcttc 2040
ctgcaacaag tcgtcttcgc ctcccttgct gtccttgccc tgctgtgtgt cctgctgctg 2100
cctgagagcc gaagccgggg gctgccccag tcactgcagg acgccgaccg cctgcgccgc 2160
tccccactcc tgcggggccg cccccgccag gaccacctgc ctctgctgcc gccctccaac 2220
tcctactggg ccggccacac ccccgagcag cactagtcct gcctggtggc cctgggagcc 2280
aggatgggac caaagtcaag gcctggggca tggctgagta ccccagacgt ctggtccagg 2340
gcagacacat tcctctcaga agcccgtgtc tcagtgcagg tggagccgtg gggacagcgt 2400
gaaggtgtct ccagccaggc cccaggcact gggaggccct gggtctcccc ccagccacac 2460
ccagtaggtg tggaggataa aggcttctgt ggaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2520
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2580
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa acaaaacaaa caacaaaaac aaaaaacaaa 2640
aaaaacacac aaaaaaaaaa aaagaaaaga agaagatcac agcgacgata taagcgtgaa 2700
tcacagagag agacaggggg acagagcgtg aactgcacac cgcagcgatg gtacgcgcag 2760
ccagctgagc gagggcgacc gtggcgcgca gcccagccga ccagcgcggt acgcgagaga 2820
gagagacgag ctagactgcc accgaaacgc agccaacaac aagagggaac acatagaagt 2880
aggtgataca gcgcacacag acaaggacga agaagaagac agcagggagg tagagcaata 2940
aacaaaaaca taaaagaaac gaaaaaaata acaaaaaaaa aaaaaaacta taaactaata 3000
caatagataa aacacattaa caatctaacc aataaaaaat aaagcgccat aaaacatatt 3060
aagcaatata taagataaaa tcgcacacaa ttaactaatc cctacatact acagaatcca 3120
acaaacatat aaacaataca tatat 3145

<210> 29
<211> 763
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1700438CB1

<400> 29
ttctcgctct gtcaccaggc tggagtgcag tggcgcgatc tcgactcact gcaagctccg 60
cctcccggat tcacgccatt gtcccgcctc agcctctcga gtagctggga atacaggcgc 120
ccaccaccat gcccggctga tgttttgtat ttttagtaga gacggggttt caccgtgtta 180
gccaggatgg cctcgatctc ctgacctcgt gatccgtctg cctcggcctc ccaaagtgct 240
gggattatag gcatgagcca ctgcgcccgg cctccttttc acctttaaaa catgcaaagg 300
ctgctctgca tgtgccatga gggatgggaa gcctactgca ggcagatggt cttcttagct 360
ggcttgtgct tggtcttcct ctacatgaca gttctggggt ctggcggcat catcactggc 420
tatgcctgta cccaggggct tggagactcc ctgcttagca tcctcacggc cctttcagct 480
ctctctggcc tgatgggcac cgttctcttc acccagctta gggggcacta tggcttggtc 540
acaactggag tcatatctag ccagctccat ttaggctgtc taatgctctg catgttttct 600

gtcttggctc ctggaaattc ttttgatctg gctgttttct cacttccatt aagtaaaaat 660
ccttcaaact atgagttatt ggtccagtgg atggaagaac agtccagagg aatggcttgg 720
ttcaggtttc tttcaaaggg ataaacatgg gtctggttct ctt 763

<210> 30
<211> 2720
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 535939CB1

<400> 30
tgtgaattat aggcagctgt gatgtcgcgt gcgcgcgttc gttagctcgg tcgttcggtg 60
cgagggaaag atgcggcgaa tcaggaaccc cagagcgcgg tggctagacc gggctccgcc 120
gcctccccca cagccccttt cctaatcgtt cagacggagc ctggtcgact tcgccggaga 180
ctgccagatc tcgttcctct tccctgtgtc atcttcttaa ttataaataa tgggggatga 240
agataaaaga attacatatg aagattcaga accatccaca ggaatgaatt acacgccctc 300
catgcatcaa gaagcacagg aggagacagt tatgaagctc aaaggtatag atgcaaatga 360
accaacagaa ggaagtattc ttttgaaaag cagtgaaaaa aagctacaag aaacaccaac 420
tgaagcaaat cacgtacaaa gactgagaca aatgctggct tgccctccac atggtttact 480
ggacagggtc ataacaaatg ttaccatcat tgttcttctg tgggctgtag tttggtcaat 540
tactggcagt gaatgtcttc ctggaggaaa cctatttgga attataatcc tattctattg 600
tgccatcatt ggtggtaaac ttttggggct tattaagtta cctacattgc ctccactgcc 660
ttctcttctt ggcatgctgc ttgcagggtt tctcatcaga aatatcccag tcatcaacga 720
taatgtgcag atcaagcaca agtggtcttc ctctttgaga agcatagccc tgtctatcat 780
tctggttcgt gctggccttg gtctggattc aaaggccctg aagaagttaa agggcgtttg 840
tgtaagactg tccatgggtc cctgtattgt ggaggcgtgc acatctgctc ttcttgccca 900
ttacctgctg ggtttaccat ggcaatgggg atttatactg ggttttgttt taggtgctgt 960
atctccagct gttgtggtgc cttcaatgct ccttttgcag ggaggaggct atggtgttga 1020
gaagggtgtc ccaaccttgc tcatggcagc tggcagcttc gatgacattc tggccatcac 1080
tggcttcaac acatgcttgg gcatagcctt ttccacaggc tctactgtct ttaatgtcct 1140
cagaggagtt ttggaggtgg taattggtgt ggcaactgga tctgttcttg gatttttcat 1200
tcagtacttt ccaagccgtg accaggacaa acttgtgtgt aagagaacat tccttgtgtt 1260
ggggttgtct gtgctagctg tgttcagcag tgtgcatttt ggtttccctg gatcaggagg 1320
actgtgcacg ttggtcatgg ctttccttgc aggcatggga tggaccagcg aaaaggcaga 1380
ggttgaaaag ataattgcag ttgcctggga catttttcag ccccttcttt ttggactaat 1440
tggagcagag gtatctattg catctctcag accagaaact gtaggccttt gtgttgccac 1500
cgtaggcatt gcagtattga tacgaatttt gactacattt ctgatggtgt gttttgctgg 1560
ttttaactta aaagaaaaga tatttatttc ttttgcatgg cttccaaagg ccacagttca 1620
ggctgcaata ggatctgtgg ctttggacac agcaaggtca catggagaga aacaattaga 1680
agactatgga atggatgtgt tgacagtggc atttttgtcc atcctcatca cagccccaat 1740
tggaagtctg cttattggtt tactgggccc caggcttctg cagaaagttg aacatcaaaa 1800
taaagatgaa gaagttcaag gagagacttc tgtgcaagtt tagaggtgaa aagagagagt 1860
gctgaacata atgtttagaa agctgctact ttttttcaaga tgcatattga aatatgtaat 1920
gtttaagctt aaaatgtaat agaaccaaaa gtgtagctgt ttctttaaac agcatttta 1980
gcccttgctc tttccatgtg ggtggtaatg atctatatca ccaaccttaa tctctctgcc 2040
tttttttca aacacccctt catcatccat cttaatttgc ataaggacat atctacttta 2100
atgtactacc acagtttaca gttaatgtgg gaaagaccag cttcagtatc ctcttcagct 2160
aggattgccc taacttttaa ctttcacagt ttcctgattc atatttgccc aggctctgat 2220
gccttgaatt ggttttggct ctctttttg gatctgtttt tgttgttaaa catcataatg 2280
cagtctctca ttaatttta ccatcattta ccctgataat ctgcctcttc tccatttctc 2340
cttcccttac taccttttctt tgaattactg taactgattg gtcccaccaa aattttaaag 2400
tacatgaagt atcttcattg gttcatcctc ttgccccctc cagatgtcaa aaaactttat 2460
cctgcccccc agctgaccac ccaggttcct ttatttcagt ggcccatgtg agtctacctt 2520

ccctaagga gtgccctaat ccagcccttt ttttgtttct tatgacccat atctttaggc 2580
tcttcccatt tctaggtggg agataggtaa gtttcaaatc tatgccagtc ttatgaatat 2640
tacattaggg taatgtgcta taatgaagaa ataaaaaata cagtgcttaa aagaaaataa 2700
aattctattt ctgtctaatg 2720

<210> 31
<211> 4464
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55118067CB1

<400> 31
ggcagctgag ttgggctgag gtgtccctag ctggctctgc ggctcttccg ggtctgggct 60
cggagattca caggcggccc gcgaggccga gcgagggacg catggccctg aggcggccgc 120
agggcttggc ggggtccgga ggttgacctc gcccccgcag ccggccttcg aggctgcctc 180
ctccaggcag cctctggggc ccgcgcccgc gcctgctcag gctcccgtgt tcaggctgcc 240
catcccctcc ccaccggcgt cccggacgtt gggacctgtg accgtggcct cgggctgggc 300
ttccaaagcc ggccgcagcc cggcgacccc cgaggcctct cgccccgggc ccctagacct 360
ctcactatga ccgcggccgc cgcctccaac tgggggctga tcacgaacat cgtgaacagc 420
atcgtagggg tcagtgtcct caccatgccc ttctgcttca aacagtgcgg catcgtcctg 480
ggggcgctgc tcttggtctt ctgctcatgg atgacgcacc agtcgtgcat gttcttggtg 540
aagtcggcca gcctgagcaa gcggaggacc tacgccggcc tggcattcca cgcctacggg 600
aaggcaggca agatgctggt ggagaccagc atgatcgggc tgatgctggg cacctgcatc 660
gccttctacg tcgtgatcgg cgacttgggg tccaacttct ttgcccggct gttcgggttt 720
caggtgggcg gcaccttccg catgttcctg ctgttcgccg tgtcgctgtg catcgtgctc 780
ccgctcagcc tgcagcggaa catgatggcc tccatccagt ccttcagcgc catggccctc 840
ctcttctaca ccgtgttcat gttcgtgatc gtgctctcct ctctcaagca cggcctcttc 900
agtgggcagt ggctgcggcg ggtcagctac gtccgctggg agggcgtctt ccgctgcatc 960
cccatcttcg gcatgtcctt cgcctgccag tcccaggtgc tgcccaccta cgacagcctg 1020
gatgagccgt cagtgaaaac catgagctcc atatttgctt cctcccttaa tgtggtcacc 1080
accttctacg tcatggtggg gtttttcggc tacgtcagct tcaccgaggc cacggccggc 1140
aacgtgctca tgcactttcc ctccaacctg gtgacggaga tgctccgtgt gggcttcatg 1200
atgtcagtgg ctgtgggctt ccccatgatg atcctgccat gcaggcaggc cctgagcacg 1260
ctgctgtgtg agcagcagca aaaagatggc acctttgcag caggggggcta catgcccct 1320
ctccggttta aagcacttac cctctctgtg gtgtttggaa ccatggttgg tggcatcctt 1380
atccccaacg tggagaccat cctgggcctc acaggagcga ccatgggaag cctcatctgc 1440
ttcatctgcc cggcgctgat ctacaagaaa atccacaaga acgcactttc ctcccaggtg 1500
gtgctgtggg tcggcctggg cgtcctggtg gtgagcactg tcaccacact gtctgtgagc 1560
gaggaggtcc ccgaggactt ggcagaggaa gcccctggcg gccggcttgg agaggccgag 1620
ggtttgatga aggtggaggc agcgcggctc tcagcccagg atccggttgt ggccgtggct 1680
gaggatggcc gggagaagcc gaagctgccg aaggagagag aggagctgga gcaggcccag 1740
atcaaggggc ccgtggatgt gcctggacgg gaagatggca aggaggcacc ggaggaggca 1800
cagctcgatc gccctgggca agggattgct gtgcctgtgg gcgaggccca ccgccacgag 1860
cctcctgttc ctcacgacaa ggtggtggta gatgaaggcc aagaccgaga ggtgccagaa 1920
gagaacaaac ctccatccag acacgcgggc ggaaaggctc caggggtcca gggccagatg 1980
gcgccgcctc tgcccgactc agaaagagag aaacaagagc cggagcaggg agaggttggg 2040
aagaggcctg gacaggccca ggccttggag gaggcgggtg atcttcctga agatccccag 2100
aaagttccag aagcagatgg tcagccagct gtccagcctg caaaggagga cctggggcca 2160
ggagacaggg gcctgcatcc tcggccccag gcagtgctgt ctgagcagca gaacggcctg 2220
gcggtgggtg gaggggaaaa ggccaagggg ggaccgccgc caggcaacgc cgccggggac 2280
acagggcagc ccgcagagga cagcgaccac ggtgggaagc ctcccctccc agcggagaag 2340
ccggctccag ggcctgggct gccgcccgag cctcgcgagc agagggacgt ggagcgagcg 2400
ggtggaaacc aggcggccag ccagctggag gaagctggca gggcggagat gctggaccac 2460

gccgtcctgc ttcaggtgat caaagaacag caggtgcagc aaaagcgctt gctggaccag 2520
caggagaagc tgctggcggt gatcgaggag cagcacaagg agatccacca gcagaggcag 2580
gaggacgagg aggataaacc caggcaggtg gaggtgcatc aagagcccgg ggcagcggtg 2640
cccagaggcc aggaggcccc tgaaggcaag gccagggaga cggtggagaa tctgcctccc 2700
ctgcctttgg accctgtcct cagagctcct gggggccgcc ctgctccatc ccaggacctt 2760
aaccagcgct ccctggagca ctctgagggg cctgtgggca gagaccctgc tggccctcct 2820
gacggcggcc ctgacacaga gcctcgggca gcccaggcca agctgagaga tggccagaag 2880
gatgccgccc ccagggcagc tggcactgtg aaggagctcc ccaagggccc ggagcaggtg 2940
cccgtgccag accccgccag ggaagccggg ggcccagagg agcgcctcgc agaggaattc 3000
cctgggcaaa gtcaggacgt tactggcggt tcccaagaca ggaaaaaacc tgggaaggag 3060
gtggcagcca ctggcaccag cattctgaag gaagccaact ggctcgtggc agggccagga 3120
gcagagacgg gggaccctcg catgaagccc aagcaagtga gccgagacct gggccttgca 3180
gcggacctgc ccggtggggc ggaaggagca gctgcacagc cccaggctgt gttacgccag 3240
ccggaactgc gggtcatctc tgatggcgag cagggtggac agcagggcca ccggctggac 3300
catggcggtc acctggagat gagaaaggcc cgcgggggggg accatgtgcc tgtgtcccac 3360
gagcagccga gaggcgggga ggacgctgct gtccaggagc ccaggcagag gccagagcca 3420
gagctggggc tcaaacgagc tgtcccgggg ggccagaggc cggacaatgc caagcccaac 3480
cgggacctga aactgcaggc tggctccgac ctccggaggc gacggcggga ccttggccct 3540
catgcagagg gtcagctggc cccgagggat ggggtcatca ttggccttaa ccccctgcct 3600
gatgtccagg tgaacgacct ccgtggcgcc ctggatgccc agctccgcca ggctgcgggg 3660
ggagctctgc aggtggtcca cagccggcag cttagacagg cgcctgggcc tccagaggag 3720
tcctagcacc tgctggccat gagggccacg ccagccactg ccctcctcgg ccagcagcag 3780
gtctgtctca gccgcatccc agccaaactc tggaggtcac actcgcctct ccccagggtt 3840
tcatgtctga ggccctcacc aagtgtgagt gacagtataa aagattcact gtggcatcgt 3900
ttccagaatg ttcttgctgt cgttctgttg cagctcttag tctgaggtcc tctgacctct 3960
agactctgag ctcactccag cctgtgagga gaaacggcct ccgctgcgag ctggctggtg 4020
cactcccagg ctcaggctgg ggagctgctg cgtctgtggt caggcctcct gctcctgcca 4080
gggagcacgc gtggtcttcg ggttgagctc ggccgtgcgt ggaggtgcgc atggctgctc 4140
atggtcccaa cacaggctac tgtgagagcc agcatccaac cccacgcttg cagtgactca 4200
gaatgataat tattatgact gtttatcgat gcttcccaca gtgtggtaga aagtcttgaa 4260
taaacacttt tgccttcacc cagcctcggt ggatgctgtt tggtgtccag gaagcacagg 4320
gagcaggggc cagacaagcc gggtgtccag gggcactggc cggtgccgcc ctttgcactc 4380
ctcatggttg gcccagccct ccattgtctg tgtttttcaa aatccaattg tggctttttt 4440
taaaaagtaa aaaacaccac tgtg 4464

<210> 32
<211> 3135
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7502087CB1

<220>
<221> unsure
<222> 2961
<223> a, t, c, g, or other

<400> 32
ggcggagacg ccgggagcca gtggcgcctg tggctccggg caggggccgc ggccgaaaga 60
tgccggtccg caggggccac gtcgctcccc aaaacactta cctggacacc atcatccgca 120
agttcgaggg ccaaagtcgg aagttcctga ttgccaatgc tcagatggag aactgcgcca 180
tcatttactg caacgacggc ttctgcgaac tcttcggcta ctcccgagtg gaggtgatgc 240
agcaaccctg cacctgcgac ttcctcacag gccccaacac accaagcagc gccgtgtccc 300
gcctagcgca ggccctgctg ggggctgagg agtgcaaggt ggacatcctc tactaccgca 360

aggatgcctc cagcttccgc tgcctggtag atgtggtgcc cgtgaagaac gaggacgggg 420
ctgtcatcat gttcattctc aacttcgagg acctggccca gctcctggcc aagtgcagca 480
gccgcagctt gtcccagcgc ctgttgtccc agagcttcct gggctccgag ggctctcatg 540
gcaggccagg cggaccaggg ccaggcacag gcaggggcaa gtacaggacc atcagccaga 600
tcccacagtt cacgctcaac ttcgtggagt tcaacttgga gaagcaccgc tccagctcca 660
ccacggagat tgagatcatc gcgccccata aggtggtgga gcggacacag aacgtcactg 720
agaaggtcac ccaggtcctg tccctgggcg cggatgtgct gccggagtac aagctgcagg 780
cgccgcgcat ccaccgctgg accatcctgc actacagccc cttcaaggcc gtgtgggact 840
ggctcatcct gctgctggtc atctacacgg ctgtcttcac gccctactca gccgccttcc 900
tgctcagcga ccaggacgaa tcacggcgtg gggcctgcag ctatacctgc agtcccctca 960
ctgtggtgga tctcatcgtg gacatcatgt tcgtcgtgga catcgtcatc aacttccgca 1020
ccacctatgt caacaccaat gatgaggtgg tcagccaccc ccgccgcatc gccgtccact 1080
acttcaaggg ctggttcctc attgacatgg tggccgccat ccctttcgac ctcctgatct 1140
tccgcactgg ctccgatgag accacaaccc tgattgggct attgaagaca gcgcggctgc 1200
tgcggctggt gcgcgtagca cggaagctgg actgctactc tgagtatggg gcggctgtgc 1260
tcttcttgct catgtgcacc ttcgcgctca tagcgcactg gctggcctgc atctggtacg 1320
ccatcggcaa tgtggagcgg ccctacctag aacacaagat cggctggctg gacagcctgg 1380
gtgtgcagct tggcaagcgc tacaacggca gcgacccagc ctcgggcccc tcggtgcagg 1440
acaagtatgt cacagccctc tacttcacct tcagcagcct caccagcgtg ggcttcggca 1500
atgtctcgcc caacaccaac tccgagaagg tcttctccat ctgcgtcatg ctcatcggct 1560
ccctgatgta cgccagcatc ttcgggaacg tgtccgcgat catccagcgc ctgtactcgg 1620
gcaccgcgcg ctaccacacg cagatgctgc gtgtcaagga gttcatccgc ttccaccaga 1680
tccccaaccc actgcgccag cgcctggagg agtatttcca gcacgcctgg tcctacacca 1740
atggcattga catgaacgcg gtgctgaagg gcttccccga gtgcctgcag gctgacatct 1800
gcctgcacct gcaccgcgca ctgctgcagc actgcccagc tttcagcggc gccggcaagg 1860
gctgcctgcg cgcgctagcc gtcaagttca agaccaccca cgcgccgcct ggggacacgc 1920
tggtgcacct cggcgacgtg ctctccaccc tctacttcat ctcccgaggc tccatcgaga 1980
tcctgcgcga cgacgtggtc gtggccatcc taggaaagaa tgacatcttt ggggaacccg 2040
tcagcctcca tgcccagcca ggcaagtcca gtgcagacgt gcgggctctg acctactgcg 2100
acctgcacaa gatccagcgg gcagatctgc tggaggtgct ggacatgtac ccggcctttg 2160
cggagagctt ctggagtaag ctggaggtca ccttcaacct gcgggacgct cctggcagcc 2220
aagaccacca aggtttcttt ctcagtgaca accagtcaga tgcagcccct cccctgagca 2280
tctcagatgc atctggcctc tggcctgagc tactgcagga aatgccccca aggcacagcc 2340
cccaaagccc tcaggaagac ccagattgct ggcctctgaa gctgggctcc aggctagagc 2400
agctccaggc ccagatgaac aggctggagt cccgcgtgtc ctcagacctc agccgcatct 2460
tgcagctcct ccagaagccc atgccccagg gccacgccag ctacattctg gaagcccctg 2520
cctccaatga cctggccttg gttcctatag cctcggagac gacgagtcca gggcccaggc 2580
tgccccaggg ctttctgcct cctgcacaga ccccaagcta tggagacttg gatgactgta 2640
gtccaaagca caggaactcc tcccccagga tgcctcacct ggctgtggca atggacaaaa 2700
ctctggcacc atcctcagaa caggaacagc ctgaggggct ctggccaccc ctagcctcac 2760
ctctacatcc cctggaagta caaggactca tctgtggtcc ctgcttctcc tccctccctg 2820
aacaccttgg ctctgttccc aagcagctgg acttccagag acatggctca gatcctggat 2880
ttgcagggag ttggggccac tgaactccaa gataaagaca ccatgagggg actgaaggtg 2940
ggcaagggga tttcctttag nctgggcatg gtggcgggcg cctgtaatcc cagctactca 3000
ggaggctgaa gcaagagaat cacttgaacc ctaaaggcag aggttgcagt gagccgagat 3060
agtgccactg cactacagcc cgggcgacag agtgagactc catctcaaaa ataaaataca 3120
attaacaaaa aaaga 3135

<210> 33
<211> 843
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500819CB1

<400> 33
ctcctcggga ctcggcgggt cctcctggga gtctcggagg ggaccggctg tgcagacgcc 60
atggagttgg tgctggtctt cctctgcagc ctgctggccc ccatggtcct ggccagtgca 120
gctgaaaagg agaaggaaat ggacccttt cattatgatt accagaccct gaggattggg 180
ggactggtgt tcgctgtggt cctcttctcg gttgggatcc tccttatcct aagtcgcagg 240
tgcaagtgca gcttctactc tgcccctggg gaatgtgtcc cctgcatatc ttctcagcaa 300
taactccatg ggctctggga ccctaccct tccaaccttc cctgcttctg agacttcaat 360
ctacagccca gctcatccag atgcagacta cagtccctgc aattgggtct ctggcaggca 420
atagttgaag gactcctgtt ccgttggggc cagcacaccg ggatggatgg agggagagca 480
gaggcctttg cttctctgcc tacgtccct tagatgggca gcagaggcaa ctcccgcatc 540
ctttgctctg cctgtcagtg gtcagagcgg tgagcgaggt gggttggaga ctcagcaggc 600
tccgtgcagc ccttgggaac agtgagaggt tgaaggtcat aacgagagtg ggaactcaac 660
ccagatcccg cccctcctgt cctctgtgtt cccgcggaaa ccaaccaaac cgtgcgctgt 720
gacccattgc tgttctctgt atcgtgatct atcctcaaca acaacagaaa aaaggaataa 780
aatatccttt gtttcctaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 840
aaa 843

<210> 34
<211> 3159
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503413CB1

<400> 34
gcgcgatggc ctcggcgctg agctatgtct ccaagttcaa gtccttcgtg atcttgttcg 60
tcaccccgct cctgctgctg ccactcgtca ttctgatgcc cgccaagttt gtcaggtgtg 120
cctacgtcat catcctcatg gccatttact ggtgcacaga agtcatccct ctggctgtca 180
cctctctcat gcctgtcttg cttttcccac tcttccagat tctggactcc aggcaggtgt 240
gtgtccagta catgaaggac accaacatgc tgttcctggg cggcctcatc gtggccgtgg 300
ctgtggagcg ctggaacctg cacaagagga tcgccctgcg cacgctcctc tgggtggggg 360
ccaagcctgc acggctgatg ctgggcttca tgggcgtcac agccctcctg tccatgtgga 420
tcagtaacac ggcaaccacg gccatgatgg tgcccatcgt ggaggccata ttgcagcaga 480
tggaagccac aagcgcagcc accgaggccg gcctggagct ggtggacaag ggcaaggcca 540
aggagctgcc agggagtcaa gtgatttttg aaggccccac tctggggcag caggaagacc 600
aagagcggaa gaggttgtgt aaggccatga ccctgtgcat ctgctacgcg gccagcatcg 660
ggggcaccgc caccctgacc gggacgggac ccaacgtggt gctcctgggc cagatgaacg 720
agttgtttcc tgacagcaag gacctcgtga actttgcttc ctggtttgca tttgcctttc 780
ccaacatgct ggtgatgctg ctgttcgcct ggctgtggct ccagtttgtt tacatgagat 840
tcaattttaa aaagtcctgg ggctgcgggc tagagagcaa gaaaaacgag aaggctgccc 900
tcaaggtgct gcaggaggag taccggaagc tggggccctt gtccttcgcg gagatcaacg 960
tgctgatctg cttcttcctg ctggtcatcc tgtggttctc ccgagacccc ggcttcatgc 1020
ccggctggct gactgttgcc tgggtggagg aaaggaaaac tccattttat ccccctcccc 1080
tgctggattg gaaggtaacc caggagaaag tgccctgggg catcgtgctg ctactagggg 1140
gcggatttgc tctggctaaa ggatccgagg cctcggggct gtccgtgtgg atggggaagc 1200
agatggagcc cttgcacgca gtgcccccgg cagccatcac cttgatcttg tccttgctcg 1260
ttgccgtgtt cactgagtgc acaagcaacg tggccaccac caccttgttc ctgcccatct 1320
ttgcctccat gtctcgctcc atcggcctca atccgctgta catcatgctg ccctgtaccc 1380
tgagtgcctc ctttgccttc atgttgcctg tggccacccc tccaaatgcc atcgtgttca 1440
cctatgggca cctcaaggtt gctgacatgg tgaaaacagg agtcataatg aacataattg 1500
gagtcttctg tgtgtttttg gctgtcaaca cctggggacg ggccatattt gacttggatc 1560
atttccctga ctgggctaat gtgacacata ttgagactta ggaagagcca caagaccaca 1620
cacacagccc ttaccctcct caggactacc gaaccttctg gcacaccttg tacagagttt 1680
tggggttcac accccaaaat gacccaacga tgtccacaca ccaccaaaac ccagccaatg 1740

ggccacctct tcctccaagc ccagatgcag agatggtcat gggcagctgg agggtaggct 1800
cagaaatgaa gggaacccct cagtgggctg ctggacccat ctttcccaag ccttgccatt 1860
atctctgtga gggaggccag gtagccgagg gatcaggatg caggctgctg tacccgctct 1920
gcctcaagca tcccccacac agggctctgg ttttcactcg cttcgtccta gatagtttaa 1980
atgggaatca gatcccctgg ttgagagcta agacaaccac ctaccagtgc ccatgtccct 2040
tccagctcac cttgagcagc ctcagatcat ctctgtcact ctggaaggga caccccagcc 2100
agggacggaa tgcctggtct tgagcaacct cccactgctg gagtgcgagt gggaatcaga 2160
gcctcctgaa gcctctggga actcctcctg tggccaccac caaaggatga ggaatctgag 2220
ttgccaactt caggacgaca cctggcttgc cacccacagt gcaccacagg ccaacctacg 2280
cccttcatca cttggttctg ttttaatcga ctggcccccc gtcccacctc tccagtgagc 2340
ctccttcaac tccttggtcc cctgttgtct gggtcaacat ttgccgagac gccttggctg 2400
gcaccctctg gggtcccccc tttctcccag gcaggtcatc ttttctggga gatgcttccc 2460
ctgccatccc caaatagcta ggatcacact ccaagtatgg gcagtgatgg cgctctgggg 2520
gccacagtgg gctatctagg ccctccctca cctgaggccc agagtggaca cagctgttaa 2580
tttccactgg ctatgccact tcagagtctt tcatgccagc gtttgagctc ctctgggtaa 2640
aatcttccct ttgttgactg gccttcacag ccatggctgg tgacaacaga ggatcgttga 2700
gattgagcag cgcttggtga tctctcagca aacaacccct gcccgtgggc caatctactt 2760
gaagttactc ggacaaagac cccaaagtgg ggcaacaact ccagagaggc tgtgggaatc 2820
ttcagaagcc cccctgtaag agacagacat gagagacaag catcttcttt cccccgcaag 2880
tccattttat ttccttcttg tgctgctctg gaagagaggc agtagcaaag agatgagctc 2940
ctggatggca ttttccaggg caggagaaag tatgagagcc tcaggaaacc ccatcaagga 3000
ccgagtatgt gtctggttcc ttgggtggga cgattcctga ccacactgtc cagctcttgc 3060
tctcattaaa tgctctgtct cccgcggaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 3120
aaaaaaaaaa aaattctcgg acgcaaggaa ttcagctgg 3159

<210> 35
<211> 1883
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500007CB1

<400> 35
ccacgcgtcc gtaaggtggg atggatagca gggtctcagg cacaaccagt aatggagaga 60
caaaaccagt gtatccagtc atggaaaaga aggaggaaga tggcaccctg gagcgggggc 120
actggaacaa caagatggag tttgtgctgt cagtggctgg ggagatcatt ggcttaggca 180
acgtctggag gtttccctat ctctgctaca aaaatggggg agaacactgt atggagttcc 240
agaagaccaa cggctccctg aatggtacct ctgagaatgc cacctctcct gtcatcgagt 300
tctgggagcg gcgggtcttg aagatctctg atgggatcca gcacctgggg gccctgcgct 360
gggagctggc tctgtgcctc ctgctggcct gggtcatctg ctacttctgc atctggaagg 420
gggtgaagtc cacaggcaag gtggtgtact tcacggccac atttccttac ctcatgctgg 480
tggtcctgtt aattcgaggg gtgacgttgc ctggggcagc ccaaggaatt cagttttacc 540
tgtacccaaa cctcacgcgt ctgtgggatc cccaggtgtg gatggatgca ggcacccaga 600
tattcttctc cttcgccatc tgtcttgggt gcctgacagc cctgggcagc tacaacaagt 660
accacaacaa ctgctacagg gactgcatcg ccctctgctt cctcaacagc ggcaccagct 720
ttgtggccgg ctttgccatc ttctccatcc tgggcttcat gtctcaggag caggggggtgc 780
ccatttctga ggtggccgag tcaggccctg gcctggcttt catcgcttac ccgcgggctg 840
tggtgatgct gcccttctct cctctctggg cctgctgttt cttcttcatg gtcgttctcc 900
tgggactgga tagccagttt gtgtgtgtag aaagcctggt gacagcgctg gtggacatgt 960
accctcacgt gttccgcaag aagaaccgga gggaagtcct catccttgga gtatctgtcg 1020
tctccttcct tgtggggctg atcatgctca cagagggcgg aatgtacgtg ttccagctct 1080
ttgactacta tgcggccagt ggcatgtgcc tcctgttcgt ggccatcttc gagtccctct 1140
gtgtggcttg ggtttacgga gccaagcgct tctacgacaa catcgaagac atgattgggt 1200
acaggccatg gcctcttatc aaatactgtt ggctcttcct cacaccagct gtgtgcacag 1260

ccacctttct cttctccctg ataaagtaca ctccgctgac ctacaacaag aagtacacgt 1320
acccgtggtg gggcgatgcc ctgggctggc tcctggctct gtcctccatg gtctgcattc 1380
ctgcctggag cctctacaga ctcggaaccc tcaagggccc cttcagagag agaatccgtc 1440
agctcatgtg cccagccgag gacctgcccc agcggaaccc agcaggaccc tcggctcccg 1500
ccacccccag gacctcactg ctcagactca cagagctaga gtctcactgc tagggggcag 1560
gcccttggat ggtgcctgtg tgcctggcct tggggatggc tgtggaggga acgtggcaga 1620
agcagcccca tgtgcttccc tgcccccgac ctggagtgga taagacaaga ggggtatttt 1680
ggagtccacc tgctgagctg gaggcctccc actgcaactt ttcagctcag gggttgttga 1740
acagatgtga aaggccagtg ccaagagtgt ccctctgaga cccttgggaa gctgggtggg 1800
ggctggtagg tggggcgaga cttgctggct tcgggccctc tcatccttca ttccattaaa 1860
tccacattct tcccgctgaa aaa 1883

<210> 36
<211> 2746
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7500025CB1

<400> 36
cgcctgtggc tccgggcagg ggccgcggcc gaaagatgtc ggtccgcagg ggccacgtcg 60
ctccccaaaa cacttacctg gacaccatca tccgcaagtt cgagggccaa agtcggaagt 120
tcctgattgc caatgctcag atggagaact gcgccatcat ttactgcaac gacggcttct 180
gcgaactctt cggctactcc cgagtggagg tgatgcagca accctgcacc tgcgacttcc 240
tcacaggccc caacacacca agcagcgccg tgtcccgcct agcgcaggcc ctgctggggg 300
ctgaggagtg caaggtggac atcctctact accgcaagga tgcctccagc ttccgctgcc 360
tggtagatgt ggtgcccgtg aagaacgagg acggggctgt catcatgtcc attctcaact 420
tcgaggacct ggcccagctc ctggccaagt gcagcagccg cagcttgtcc cagcgcctgt 480
tgtcccagag cttcctgggc tccgagggct ctcatggcag gccaggcgga ccaggcccag 540
gcacaggcag gggcaagtac aggaccatca gccagatccc acagttcacg ctcaacttcg 600
tggagttcaa cttggagaag caccgctcca gctccaccac ggagattgag atcatcgcgc 660
cccataaggt ggtggagcgg acacagaacg tcactgagaa ggtcacccag gtcctgtccc 720
tgggcgcaga tgtgctgccg gagtacaagc tgcaggcgcc gcgcatccac cgctggacca 780
tcctgcacta cagccccttc aaggccgtgt gggactggct catcccgctg ctggtcatct 840
acacggctgt cttcacgccc tactcagccg ccttcctgct cagcgatcag gacgaatcac 900
ggcgtggggc ctgcagctat acctgcagtc ccctcactgt ggtggatctc atcgtggaca 960
tcatgttcgt cgtggacatc gtcatcaact tccgcaccac ctatgtcaac accaatgatg 1020
aggtggtcag ccacccccgc cgcatcgccg tccactactt caagggctgg ttcctcattg 1080
acatggtggc cgccatccct ttcgacctcc tgatcttccg cactggctcc gatgagacca 1140
caaccctgat tgggctattg aagacagcgc ggctgctgcg gctggtgcgc gtagcacgga 1200
agctggaccg ctactctgag tatggggcgg ctgtgctctt cttgctcatg tgcaccttcg 1260
cgctcatagc gcactggctg gcctgcatct gcagcctcac cagcgtgggc ttcggcaatg 1320
tctcgcccaa caccaactcc gagaaggtct tctccatctg cgtcatgctc atcggctccc 1380
tgatgtacgc cagcatcttc gggaacgtgt ccgcgatcat ccagcgcctg tactcgggca 1440
ccgcgcgcta ccacacgcag atgctgcgtg tcaaggagtt catccgcttc caccagatcc 1500
ccaacccact gcgccagcgc ctggaggagt atttccagca cgcctggtcc tacaccaatg 1560
gcattgacat gaacgcggtg ctgaagggct tccccgagtg cctgcaggct gacatctgcc 1620
tgcacctgca ccgcgcactg ctgcagcact gcccagcttt cagcggcgcc ggcaagggct 1680
gcctgcgcgc gctagccgtc aagttcaaga ccacccacgc gccgcctggg gacacgctgg 1740
tgcacctcgg cgacgtgctc tccaccctct acttcatctc ccgaggctcc atcgagatcc 1800
tgcgcgacga cgtggtcgtg gccatcctag gaaagaatga catctttggg gaacccgtca 1860
gcctccatgc ccagccaggc aagtccagtg cagacgtgcg ggctctgacc tactgcgacc 1920
tgcacaagat ccagcgggca gatctgctgg aggtgctgga catgtacccg gcctttgcgg 1980
agagcttctg gagtaagctg gaggtcacct tcaacctgcg ggacgctcct ggcagccaag 2040

accaccaagg tttctttctc agtgacaacc agtcagatgc agcccctccc ctgagcatct 2100
cagatgcatc tggcctctgg cctgagctac tgcaggaaat gcccccaagg cacagccccc 2160
aaagccctca ggaagaccca gattgctggc ctctgaagct gggctccagg ctagagcagc 2220
tccaggccca gatgaacagg ctggagtccc gcgtgtcctc agacctcagc cgcatcttgc 2280
agctcctcca gaagcccatg ccccagggcc acgccagcta cattctggaa gcccctgcct 2340
ccaatgacct ggccttggtt cctatagcct cggagacgac gagtccaggg cccaggctgc 2400
cccagggctt tctgcctcct gcacagaccc caagctatgg agacttggat gactgtagtc 2460
caaagcacag gaactcctcc cccaggatgc ctcacctggc tgtggcaatg gacaaaactc 2520
tggcaccatc ctcagaacag gaacagcctg aggggctctg gccaccccta gcctcacctc 2580
tacatccect ggaagtacaa ggactcatct gtggtccctg cttctcctcc ctccctgaac 2640
accttggctc tgttcccaag cagctggact tccagagaca tggctcagat cctggatttg 2700
cagggagttg gggccactga actccaagat aaagacacca tgaggg 2746

<210> 37
<211> 2868
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7502736CB1

<400> 37
caaaacactt acctggacac catcatccgc aagttcgagg gccaaagtcg gaagttcctg 60
attgccaatg ctcagatgga gaactgcgcc atcatttact gcaacgacgg cttctgcgaa 120
ctcttcggct actcccgagt ggagcctcca gcttccgctg cctggtagat gtggtgcccg 180
tgaagaacga ggacggggct gtcatcatgt tcattctcaa cttcgaggac ctggcccagc 240
tcctggccaa gtgcagcagc cgcagcttgt cccagcgcct gttgtcccag agcttcctgg 300
gctccgaggg ctctcatggc aggccaggcg gaccagggcc aggcacaggc aggggcaagt 360
acaggaccat cagccagatc ccacagttca cgctcaactt cgtggagttc aacttggaga 420
agcaccgctc cagctccacc acggagattg agatcatcgc gccccataag gtggtggagc 480
ggacacagaa cgtcactgag aaggtcaccc aggtcctgtc cctgggcgcg gatgtgctgc 540
cggagtacaa gctgcaggcg ccgcgcatcc accgctggac catcctgcac tacagccccc 600
ttcaaggccg tgtgggactg gctcatcctg ctgctggtca tctacacggc tgtcttcacg 660
ccctactcag ccgccttcct gctcagcgat caggacgaat cacggcgtgg ggcctgcagc 720
tatacctgca gtcccctcac tgtggtggat ctcatcgtgg acatcatgtt cgtcgtggac 780
atcgtcatca acttccgcac cacctatgtc aacaccaatg atgaggtggt cagccacccc 840
cgccgcatcg ccgtccacta cttcaagggc tggttcctca ttgacatggt ggccgccatc 900
cctttcgacc tcctgatctt ccgcactggc tccgatgaga ccacaaccct gattgggcta 960
ttgaagacag cgcggctgct gcggctggtg cgcgtagcac ggaagctgga ccgctactct 1020
gagtatgggg cggctgtgct cttcttgctc atgtgcacct tcgcgctcat agcgcactgg 1080
ctggcctgca tctggtacgc catcggcaat gtggagcggc cctacctaga acacaagatc 1140
ggctggctgg acagcctggg tgtgcagctt ggcaagcgct acaacggcag cgacccagcc 1200
tcgggcccct cggtgcagga caagtatgtc acagccctct acttcacctt cagcagcctc 1260
accagcgtgg gcttcggcaa tgtctcgccc aacaccaact ccgagaaggt cttctccatc 1320
tgcgtcatgc tcatcggctc cctgatgtac gccagcatct tcgggaacgt gtccgcgatc 1380
atccagcgcc tgtactcggg caccgcgcgc taccacacgc agatgctgcg tgtcaaggag 1440
ttcatccgct tccaccagat ccccaaccca ctgcgccagc gcctggagga gtatttccag 1500
cacgcctggt cctacaccaa tggcattgac atgaacgcgg tgctgaaggg cttccccgag 1560
tgcctgcagg ctgacatctg cctgcacctg caccgcgcac tgctgcagca ctgcccagct 1620
ttcagcggcg ccggcaaggg ctgcctgcgc gcgctagccg tcaagttcaa gaccacccac 1680
gcgccgcctg gggacacgct ggtgcacctc ggcgacgtgc tctccaccct ctacttcatc 1740
tcccgaggct ccatcgagat cctgcgcgac gacgtggtcg tggccatcct aggaaagaat 1800
gacatctttg gggaacccgt cagcctccat gcccagccag gcaagtccag tgcagacgtg 1860
cgggctctga cctactgcga cctgcacaag atccagcggg cagatctgct ggaggtgctg 1920
gacatgtacc cggcctttgc ggagagcttc tggagtaagc tggaggtcac cttcaacctg 1980

cgggacgctc ctggcagcca agaccaccaa ggtttctttc tcagtgacaa ccagtcagat 2040
gcagcccctc ccctgagcat ctcagatgca tctggcctct ggcctgagct actgcaggaa 2100
atgccccaa ggcacagccc ccaaagccct caggaagacc cagattgctg gcctctgaag 2160
ctgggctcca ggctagagca gctccaggcc cagatgaaca ggctggagtc ccgcgtgtcc 2220
tcagacctca gccgcatctt gcagctcctc cagaagccca tgccccaggg ccacgccagc 2280
tacattctgg aagcccctgc ctccaatgac ctggccttgg ttcctatagc ctcggagacg 2340
acgagtccag ggcccaggct gccccagggc tttctgcctc ctgcacagac cccaagctat 2400
ggagacttgg atgactgtag tccaaagcac aggaactcct cccccaggat gcctcacctg 2460
gctgtggcaa tggacaaaac tctggcacca tcctcagaac aggaacagcc tgaggggctc 2520
tggccacccc tagcctcacc tctacatccc ctggaagtac aaggactcat ctgtggtccc 2580
tgcttctcct ccctccctga acaccttggc tctgttccca agcagctgga cttccagaga 2640
catggctcag atcctggatt tgcagggagt tggggccact gaactccaag ataaagacac 2700
catgagggga ctgaaggtgg gcaaggggat ttcctttagc tgggcatggt ggcgggcgcc 2760
tgtaatccca gctactcagg aggctgaagc aagagaatca cttgaaccct aaaggcagag 2820
gttgcagtga gccgagatag tgccactgca ctacagcccg ggcgacag 2868

<210> 38
<211> 1906
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503570CB1

<400> 38
ccttaagggg cgggccgggg cggggctccg ctgcccttc ccagaggccg cgcctgctgc 60
tgagcagatg cagtagccga aactgcgcgg aggcacagag gccggggaga gcgttctggg 120
tccgagggtc caggtagggg ttgagccacc atctgaccgc aagctgcgtc gtgtcgccgg 180
ttctgcaggc accatgagcc aggacaccga ggtggatatg aaggaggtgg agctgaatga 240
gttagagccc gagaagcagc cgatgaacgc ggcgtctggg gcggccatgt ccctggcggg 300
agccgagaag aatggtctgg tgaagatcaa ggtggcggaa gacgaggcgg aggcggcagc 360
cgcggctaag ttcacgggcc tgtccaagga ggagctgctg aaggtggcag gcagccccgg 420
ctgggtacgc acccgctggg cactgctgct gctcttctgg ctcggctggc tcggcatgct 480
tgctggtgcc gtggtcataa tcgtgcgagc gccgcgttgt cgcgagctac cggcgcagaa 540
gtggtggcac acgggcgccc tctaccgcat cggcgacctt caggccttcc agggccacgg 600
cgcgggcaac ctggcgggtc tgaaggggcg tctcgattac ctgagctctc tgaaggtgaa 660
gggccttgtg ctgggtccaa ttcacaagaa ccagaaggat gatgtcgctc agactgactt 720
gctgcagatc gaccccaatt ttggctccaa ggaagatttt gacagtctct tgcaatcggc 780
taaaaaaaag agcatccgtg tcattctgga ccttactccc aactaccggg gtgagaactc 840
gtggttctcc actcaggttg acactgtggc caccaaggtg aaggatgctc tggagttttg 900
gctgcaagct ggcgtggatg ggttccaggt tcgggacata gagaatctga aggatgcatc 960
ctcattcttg gctgagtggc aaaatatcac caagggcttc agtgaagaca ggctcttgat 1020
tgcggggact aactcctccg accttcagca gatcctgagc ctactcgaat ccaacaaaga 1080
cttgctgttg actagctcat acctgtctga ttctggttct actggggagc atacaaaatc 1140
cctagtcaca cagtatttga atgccactgg caatcgctgg tgcagctgga gtttgtctca 1200
ggcaaggctc ctgacttcct tcttgccggc tcaacttctc cgactctacc agctgatgct 1260
cttcaccctg ccagggaccc ctgttttcag ctacggggat gagattggcc tggatgcagc 1320
tgcccttcct ggacagggcc agagtgaaga ccctggctcc ctcctttcct tgttccggcg 1380
gctgagtgac cagcggagta aggagcgctc cctactgcat ggggacttcc acgcgttctc 1440
cgctgggcct ggactcttct cctatatccg ccactgggac cagaatgagc gttttctggt 1500
agtgcttaac tttggggatg tgggcctctc ggctggactg caggcctccg acctgcctgc 1560
cagcgccagc ctgccagcca aggctgacct cctgctcagc acccagccag gccgtgagga 1620
gggctcccct cttgagctgg aacgcctgaa actggagcct cacgaagggc tgctgctccg 1680
cttcccctac gcggcctgac ttcagcctga catggaccca ctacccttct cctttccttc 1740
ccaggccctt tggcttctga tttttctctt ttttaaaaac aaacaaacaa actgttgcag 1800

attatgagtg aacccccaaa tagggtgttt tctgccttca aataaaagtc acccctgcat 1860
ggtgaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa attcgg 1906

<210> 39
<211> 2506
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7504008CB1

<400> 39
cggctctagc tcgagcagca ggagcagccc gcaccggaca acttgcgagc catggggctg 60
gcggatgcgt cgggaccgag ggacacacag gcactgctgt ctgcaacaca agcaatggac 120
ctgcggaggc gagactacca catggaacgg ccgctgctga accaggagca tttggaggag 180
ctggggcgct ggggctcagc acctaggacc caccagtggc ggacctggtt gcagtgctcc 240
cgtgctcggg cctatgccct tctgctccaa cacctcccgg ttttggtctg gttacccccgg 300
tatcctgtgc gtgactggct cctgggtgac ctgttatccg gcctgagtgt ggccatcatg 360
cagcttccgc agggcttggc ctacgccctc ctggctggat tgccccccgt gtttggcctc 420
tatagctcct tctaccctgt cttcatctac ttcctgtttg gcacttcccg gcacatctcc 480
gtggggacct ttgctgtcat gtctgtgatg gtgggcagtg tgacagaatc cctggccccg 540
caggccttga acgactccat gatcaatgag acagccagag atgctgcccg ggtacaggtg 600
gcctccacac tcagtgtcct ggttggcctc ttccaggtgg ggctgggcct gatccacttc 660
ggcttcgtgg tcacctacct gtcagaacct cttgtccgag gctataccac agctgcagct 720
gtgcaggtct tcgtctcaca gctcaagtat gtgtttggcc tccatctgag cagccactct 780
gggccactgt ccctcatcta tacagtgctg gaggtctgct ggaagctgcc ccagagcaag 840
ctcatcgggg ccacaggcat ctcctatggc atgggtctaa agcacagatt tgaggtagat 900
gtcgtgggca acatccctgc agggctggtg cccccagtgg cccccaacac ccagctgttc 960
tcaaagctcg tgggcagcgc cttcaccatc gctgtggttg ggtttgccat tgccatctca 1020
ctggggaaga tcttcgccct gaggcacggc taccgggtgg acagcaacca ggagctggtg 1080
gccctgggcc tcagtaacct tatcggaggc atcttccagt gcttccccgt gagttgctct 1140
atgtctcgga gcctggtaca ggagagcacc gggggcaact cgcaggttgc tggagccatc 1200
tcttcccttt tcatcctcct catcattgtc aaacttgggg aactcttcca tgacctgccc 1260
aaggcggtcc tggcagccat catcattgtg aacctgaagg gcatgctgag gcagctcagc 1320
gacatgcgct ccctctggaa ggccaatcgg gcggatctgc ttatctggct ggtgaccttc 1380
acggccacca tcttgctgaa cctggacctt ggcttggtgg ttgcggtcat cttctccctg 1440
ctgctcgtgg tggtccggac acagatgccc cactactctg tcctggggca ggtgccagac 1500
acggatattt acagagatgt ggcagagtac tcagaggcca aggaagtccg gggggtgaag 1560
gtcttccgct cctcggccac cgtgtacttt gccaatgctg agttctacag tgatgcgctg 1620
aagcagaggt gtggtgtgga tgtcgacttc ctcatctccc agaagaagaa actgctcaag 1680
aagcaggagc agctgaagct gaagcaactg cagaaagagg agaagcttcg gaaacaggct 1740
gcctccccca agggcgcctc agtttccatt aatgtcaaca ccagccttga agacatgagg 1800
agcaacaacg ttgaggactg caagatgatg caggtgagct caggagataa gatggaagat 1860
gcaacagcca atggtcaaga agactccaag gccccagatg ggtccacact gaaggccctg 1920
ggcctgcctc agccagactt ccacagcctc atcctggacc tgggtgccct ctcctttgtg 1980
gacactgtgt gcctcaagag cctgaagaat attttccatg acttccggga gattgaggtg 2040
gaggtgtaca tggcggcctg ccacagccct gtggtcagcc agcttgaggc tgggcacttc 2100
ttcgatgcat ccatcaccaa gaagcatctc tttgcctctg tccatgatgc tgtcaccttt 2160
gccctccaac acccgaggcc tgtccccgac agccctgttt cggtcaccag actctgaaca 2220
tgctacatcc tgcccaagac tgcacctctg gaggtgcagg gcacccttga gaagcccctc 2280
acccctaggc cgcctccagg tgctacccag gagtcccctc catgtacaca cacacaactc 2340
agggaaggag gtcctgggac tccaagttca gcgctccagg tctgggacag ggcctgcatg 2400
cagtcaggct ggcagtggcg cggtacaggg agggaactgg tgcatatttt agcctcagga 2460
ataaagattt gtctgctcaa aaaaaaaaaa aaaaaaaacc gcggtc 2506

<210> 40
<211> 3836
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503559CB1

<400> 40
ctggaccttt aatccactgt aggtatggac agggaagaaa ggaagaccat caatcagggt 60
caagaagatg aaatggagat ttatggttac aatttgagtc gctggaagct tgccatagtt 120
tctttaggag tgatttgctc tggtgggttt ctcctcctcc tcctctattg gatgcctgag 180
tggcgggtga aagcgacctg tgtcagagct gcaattaaag actgtgaagt agtgctgctg 240
aggactactg atgaattcaa aatgtggttt tgtgcaaaaa ttcgcgttct ttctttggaa 300
acttacccag tttcaagtcc aaaatctatg tctaataagc tttcaaatgg ccatgcagtt 360
tgtttaattg agaatcccac tgaagaaaat aggcacagga tcagtaaata ttcacagact 420
gaatcacaac agattcgtta tttcacccac catagtgtaa aatatttctg gaatgatacc 480
attcacaatt ttgatttctt aaagggactg gatgaaggtg tttcttgtac gtcaatttat 540
gaaaagcata gtgcaggact gacaaagggg atgcatgcct acagaaaact gctttatgga 600
gtaaatgaaa ttgctgtaaa agtgccttct gtttttaagc ttctaattaa agaggttctc 660
aacccatttt acattttcca gctgttcagt gttatactgt ggagcactga tgaatactat 720
tactatgctc tagctattgt ggttatgtcc atagtatcaa tcgtaagctc actatattcc 780
attagaaagc aatatgttat gttgcatgac atggtggcaa ctcatagtac cgtaagagtt 840
tcagtttgta gagtaaatga agaaatagaa gaaatctttt ctaccgacct tgtgccagga 900
gatgtcatgg tcattccatt aaatgggaca ataatgcctt gtgatgctgt gcttattaat 960
ggtacctgca ttgtaaacga aagcatgtta acaggagaaa gtgttccagt gacaaagact 1020
aatttgccaa atccttcagt ggatgtgaaa ggaataggag atgaattata taatccagaa 1080
acacataaac gacatacttt gttttgtggg acaactgtta ttcagactcg tttctacact 1140
ggagaactcg tcaaagccat agttgttaga acaggattta gtacttccaa aggacagctt 1200
gttcgttcca tattgtatcc caaaccaact gattttaaac tctacagaga tgcctacttg 1260
tttctactat gtcttgtggc agttgctggc attgggttta tctacactat tattaatagc 1320
attttaaatg aggtacaagt tggggtcata attatcgagt ctcttgatat tatcacaatt 1380
actgtgcccc ctgcacttcc tgctgcaatg actgctggta ttgtgtatgc tcagagaaga 1440
ctgaaaaaaa tcggtatttt ctgtatcagt cctcaaagaa taaatatttg tggacagctc 1500
aatcttgttt gctttgacaa gactggaact ctaactgaag atggtttaga tctttggggg 1560
attcaacgag tggaaaatgc acgatttctt tcaccagaag aaaatgtgtg caatgagatg 1620
ttggtaaaat cccagtttgt tgcttgtatg gctacttgtc attcacttac aaaaattgaa 1680
ggagtgctct ctggtgatcc acttgatctg aaaatgtttg aggctattgg atggattctg 1740
gaagaagcaa ctgaagaaga aacagcactt cataatcgaa ttatgcccac agtggttcgt 1800
cctcccaaac aactgcttcc tgaatctacc cctgcaggaa accaagaaat ggagctgttt 1860
gaacttccag ctacttatga gataggaatt gttcgccagt tcccattttc ttctgctttg 1920
caacgtatga gtgtggttgc cagggtgctg ggggatagga aaatggacgc ctacatgaaa 1980
ggagcgcccg aggccattgc cggtctctgt aaacctgaaa cagttcctgt cgattttcaa 2040
aacgttttgg aagacttcac taaacagggc ttccgtgtga ttgctcttgc acacagaaaa 2100
ttggagtcaa aactgacatg gcataaagta cagaatatta gcagagatgc aattgagaac 2160
aacatggatt ttatgggatt aattataatg cagaacaaat taaagcaaga aacccctgca 2220
gtacttgaag atttgcataa agccaacatt cgcaccgtca tggtcacagg tgacagtatg 2280
ttgactgctg tctctgtggc cagagattgt ggaatgattc tacctcagga taaagtgatt 2340
attgctgaag cattacctcc aaaggatggg aaagttgcca aaataaattg gcattatgca 2400
gactccctca cgcagtgcag tcatccatca gcaattgacc cagaggctat tccggttaaa 2460
ttggtccatg atagcttaga ggatcttcaa atgactcgtt atcatttgc aatgaatgga 2520
aaatcattct cagtgatact ggagcatttt caagaccttg ttcctaagtt gatgttgcat 2580
ggcaccgtgt ttgcccgtat ggcacctgat cagaagacac agttgataga agcattgcaa 2640
aatgttgatt attttgttgg gatgtgtggt gatggcgcaa atgattgtgg tgctttgaag 2700
agggcacacg gaggcatttc cttatcggag ctcgaagctt cagtggcatc tccctttacc 2760

tctaagactc ctagtatttc ctgtgtgcca aaccttatca gggaaggccg tgctgcttta 2820
ataacttcct tctgtgtgtt taaattcatg gcattgtaca gcattatcca gtacttcagt 2880
gttactctgc tgtattctat cttaagtaac ctaggagact tccagtttct cttcattgat 2940
ctggcaatca ttttggtagt ggtatttaca atgagtttaa atcctgcctg gaaagaactt 3000
gtggcacaaa gaccaccttc gggtcttata tctggggccc ttctcttctc cgttttgtct 3060
cagattatca tctgcattgg atttcaatct ttgggttttt tttgggtcaa acagcaacct 3120
tggtatgaag tgtggcatcc aaaatcagat gcttgtaata caacaggaag cgggttttgg 3180
aattcttcac acgtagacaa tgaaaccgaa cttgatgaac ataatataca aaattatgaa 3240
aataccacag tgtttttttat ttccagtttt cagtacctca tagtggcaat tgccttttca 3300
aaaggaaaac ccttcaggca accttgctac aaaaattatt tttttgtttt ttctgtgatt 3360
tttttatata tttttatatt attcatcatg ttgtatccag ttgcctctgt tgaccaggtt 3420
cttcagatag tgtgtgtacc atatcagtgg cgtgtaacta tgctcatcat tgttcttgtc 3480
aatgcctttg tgtctatcac agtggaggag tcagtggatc ggtggggaaa atgctgctta 3540
ccctgggccc tgggctgtag aaagaagaca ccaaaggcaa agtacatgta tctggcgcag 3600
gagctcttgg ttgatccaga atggccacca aaacctcaga caaccacaga agctaaagct 3660
ttagttaagg agaatggatc atgtcaaatc atcaccataa catagcagtg aatcagtctc 3720
agtggtattg ctgatagcag tattcaggaa tatgtgattt taggagtttc tgatcctgtg 3780
tgtcagaatg gcactagttc agtttatgtc ccttctgata tagtagctta tttgac 3836

<210> 41
<211> 2240
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6243872CB1

<400> 41
agcgggtttt gcctcctgcc ctaggagaaa cgttcgctct ggaggcttgg gcggcaagag 60
ccccttgtgg ccaccgagtc ctccgacgcc ctcggccagg ctggcctttg ggttggccca 120
ggcaggacgg gcagccgaga gcactcgggc cgcgtcgcca ggagccgccc agggtgagcc 180
atgttcgtag gcgtcgcccg gcactctggg agccaggatg aagtctcaag gggagtagag 240
ccgctggagg ccgcgcgggc ccagcctgct aaggacagga gggccaaggg aaccccgaag 300
tcctcgaagc ccgggaaaaa acaccggtat ctgagactac ttccagaggc cttgataagg 360
ttcggcggtt tccgaaaaag gaaaaaagcc aagtcctcag tttccaagaa gccgggagaa 420
gtggatgaca gtttggagca gccctgtggt ttgggctgct tagtcagcac ctgctgtgag 480
tgttgcaata acattcgctg cttcatgatt ttctactgca tcctgctcat atgtcaaggt 540.
gtggtgtttg gtcttataga tgtcagcatt ggtgattttc agaaggaata tcaactgaaa 600
accattgaga agttggcatt ggaaaagagt tacgatattt catctggcct gactgtgcag 660
ggaatagcag gaatgcctct ttatatcctt ggaataacct ttattgatga gaatgttgct 720
acacactcag ctggtatcta tttaggtatt gcagaatgta catcaatgat tggatatgct 780
ctgggttatg tgctaggagc accactagtt aaagtccctg agaatactac ttctgcaaca 840
aacactacag tcaataatgg tagtccagaa tggctatgga cttggtggat taattttctt 900
tttgccgctg tcgttgcatg gtgtacatta ataccattgt catgctttcc aaacaatatg 960
ccaggttcaa cacggataaa agctaggaaa cgtaaacagc ttcatttttt tgacagcaga 1020
cttaaagatc tgaaacttgg aattaatatc aaggatttat gtgctgctct ttggattctg 1080
atgaagaatc cagtgctcat atgcctagct ctgtcaaaag ctacagaata tttagttatt 1140
attggagctt ctgaattttt gcctatatat ttagaaaatc agtttatatt aacacccact 1200
gtggcaacta cacttgcagg acttgtttta attccaggag gtgcacttgg ccagcttctg 1260
ggaggtgtca ttgtttccac attagaaatg tcttgtaaag cccttatgag atttataatg 1320
gttacatctg tgatatcact tatactgctt gtgtttatta tttttgtacg ctgtaatcca 1380
gtgcaatttg ctgggatcaa tgaagattat gatggaacag ggaagttggg aaacctcacg 1440
gctccttgca atgaaaaatg tagatgctca tcttcaattt attcttctat atgtggaaga 1500
gatgatattg aatatttttc tccctgcttt gcaggtattg tttcctgtct tcaatactca 1560
cagatgtact acaattgttc ttgcattaaa gaaggattaa taactgcaga tgcagaaggt 1620

gattttattg atgccagacc cgggaaatgt gatgcaaagt gctataagtt acctttgttc 1680
attgctttta tcttttctac acttatattt tctggttttt ctggtgtacc aatcgtcttg 1740
gccatgacgc gggttgtacc tgacaaactg cgttctctgg ccttgggtgt aagctatgtg 1800
attttgagaa tatttgggac tattcctgga ccatcaatct ttaaaatgtc aggagaaact 1860
tcttgtattt tacgggatgt taataaatgt ggacacacag gacgttgttg gatatataac 1920
aagacaaaaa tggctttctt attggtagga atatgttttc tttgcaaact atgcactatc 1980
atcttcacta ctattgcatt tttcatatac aaacgtcgtc taaatgagaa cactgacttc 2040
ccagatgtaa ctgtgaagaa tccaaaagtt aagaaaaaag aagaaactga cttgtaactg 2100
gatcatcatt gtgattgcag atcatttgag gatcagagtg tgaaaaggag tttctctttt 2160
acagattctc caagatttgt ttctgtgccc aactttcaga agaggaaaat cacacattat 2220
gtttacataa gtagcaaaaa 2240

<210> 42
<211> 2807
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 90011608CB1

<400> 42
gggagtcaga agtcaagtga actcagcccg cctctgtgta ctttgcactt ttccatttcc 60
cttggtacca ggcactttca tacttaatcc atagtggagc tgtcacagtg agcaactctg 120
acaatgacag cttctacccc agaggcgacc ccaaacatgg agctaaaggc tccagctgca 180
ggaggtctta atgctggccc tgtcccccca gctgccttgt ccacgcagag acttcggaat 240
gaagactacc acgactacag ctccacggac gtgagccctg aggagagccc gtcggaaggc 300
ctcaacaacc tctcctcccc gggctcctac cagcgctttg gtcaaagcaa tagcacaaca 360
tggttccaga ccttgatcca cctgttaaaa ggcaacattg gcacaggact cctgggactc 420
cctctggcgg tgaaaaatgc aggcatcgtg atgggtccca tcagcctgct gatcataggc 480
atcgtggccg tgcactgcat gggtatcctg gtgaaatgtg ctcaccactt ctgccgcagg 540
ctgaataaat cctttgtgga ttatggtgat actgtgatgt atggactaga atccagcccc 600
tgctcctggc tccggaacca cgcacactgg ggaagacgtg ttgtggactt cttcctgatt 660
gtcacccagc tgggattctg ctgtgtctat tttgtgtttc tggctgacaa ctttaaacag 720
gtgatagaag cggccaatgg gaccaccaat aactgccaca acaatgagac ggtgattctg 780
acgcctacca tggactcgcg actctacatg ctctccttcc tgcccttcct ggtgctgctg 840
gttttcatca ggaacctccg agccctgtcc atcttctccc tgttggccaa catcaccatg 900
ctggtcagct tggtcatgat ctaccagttc attgttcaga ggatcccaga ccccagccac 960
ctccccttgg tggccccttg gaagacctac cctctcttct ttggcacagc gatttttca 1020
tttgaaggca ttggaatggt tctgcccctg gaaaacaaaa tgaaggatcc tcggaagttc 1080
ccactcatcc tgtacctggg catggtcatc gtcaccatcc tctacatcag cctggggtgt 1140
ctggggtacc tgcaatttgg agctaatatc caaggcagca taaccctcaa cctgcccaac 1200
tgctggttgt accagtcagt taagctgctg tactccatcg ggatcttttt cacctacgca 1260
ctccagttct acgtcccggc tgagatcatc atccccttct ttgtgtcccg agcgcccgag 1320
cactgtgagt tagtggtgga cctgtttgtg cgcacagtgc tggtctgcct gacatgcatc 1380
ttggccatcc tcatcccccg cctggacctg gtcatctccc tggtgggctc cgtgagcagc 1440
agcgccctgg ccctcatcat cccaccgctc ctggaggtca ccaccttcta ctcagagggc 1500
atgagccccc tcaccatctt taaggacgcc ctgatcagca tcctgggctt cgtgggcttt 1560
gtggtgggga cctatgaggc tctctatgag ctgatccagc caagcaatgc tcccatcttc 1620
atcaattcca cctgtgcctt catataggga tctgggttcg tctctgcagc tgcctacccc 1680
tgccccatgt gtcccccgtt acctgtcctc agagcctcag gtatggtcca ggctctgagg 1740
aaagtcaggg ttgctgtgtg ggaacccctc tgcctggcac ctggataccc tgggccaggt 1800
aacctgaggg caggggagag gtggggtggc agacacgcag aagtgctact agtgacaggg 1860
ctgccatcgc tcacctgtac ctatttacac ccagaacttt ccagctcccc ctcatcatgc 1920
ctcctccttc ctacctgcct cccctctgct ggtgcacctc gcccaactca ttcttactgc 1980
acagttcact ttatttaaca attttcatgt cccccacctc atgttttcac cttttctggg 2040

ccaggcatag attaagtaac tgggaacgcc ccctctttat aaagctgggc ttctttctca 2100
tctctctccc aaatgttgta tctcagtatt cttcctattc gagtctccag ggggtggctg 2160
gacctacctg gtcatttgaa acaggccccc aagctggagt tttttaatct ggactctctg 2220
gcttgctgtg acccctaagg caatgcttct cttccctgga ttccttagtg tgggtcacag 2280
tactgtgttc ttagttgctt tagctcttaa aacatacgaa gtgttgccta aactgaaaat 2340
atttatcttt tatttaaaat cagatttttg tttttagact gtcttagatc tggggctatt 2400
acgaatcact tcttcttcag taaactttga ctcaacttct cctgctgaaa agaagctcgc 2460
tccagatgtc tgcatgggtc ctcggcactc ttggctgagg actcaaaggt tttaatcagg 2520
atcgtctaaa aatgtacctc ggtgaggagg cacagatttt gcctcctgtt gaccagcctg 2580
gtttcatacc gaaaagacat tgaaggactg cagaaatgta tgggtgcacc gggccgaggg 2640
aagggtggct gagtgagagg catataaaat ggggctgtgt gcatgcaggc ccatgtttca 2700
gcctcagccc acgccaggtg aaaggatcag caatgctctg tcgccatcgt gctgggacga 2760
caccagctct attgccaccg atgagtagct gaggtcagtg tgcacag 2807

<210> 43
<211> 3201
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 90024583CB1

<400> 43
atggcctcgg cgctgagcta tgtctccaag ttcaagtcct tcgtgatctt gttcgtcacc 60
ccgctcctgc tgctgccact cgtcattctg atgcccgcca agtttgtcag gtgtgcctac 120
gtcatcatcc tcatggccat ttactggtgc acagaagtca tccctctggc tgtcacctct 180
ctcatgcctg tcttgctttt cccactcttc cagattctgg actccaggca ggtatgtgtc 240
cagtacatga aggacaccaa catgctgttc ctgggcggcc tcatcgtggc cgtggctgtg 300
gagcgctgga acctgcacaa gaggatcgcc ctgcgcacgc tcctctgggt gggggccaag 360
cctgcacggc tgatgctggg cttcatgggc gtcacagccc ccctgtccat gtggatcagt 420
aacacggcaa ccacggccat gatggtgccc atcgtggagg ccatattgca gcagatggaa 480
gccacaagcg cagccaccga ggccggcctg gagctggtgg acaagggcaa ggccaaggag 540
ctgccaggga gtcaagtgat ttttgaaggc cccactctgg ggcagcagga agaccaagag 600
cggaagaggt tgtgtaaggc catgaccctg tgcatctgct acgcggccag catcgggggc 660
accgccaccc tgaccgggac gggacccaac gtggtgctcc tgggccagat gaacgagttg 720
tttcctgaca gcaaggacct cgtgaacttt gcttcctggt ttgcatttgc ctttcccaac 780
atgctggtga tgctgctgtt cgcctggctg tggctccagt ttgtttacat gagattcaat 840
tttaaaaagt cctggggctg cgggctagag agcaagaaaa acgagaaggc tgccctcaag 900
gtgctgcagg aggagtaccg gaagctgggg cccttgtcct tcgcggagat caacgtgctg 960
atctgcttct tcctgctggt catcctgtgg ttctcccgag accccggctt catgcccggc 1020
tggctgactg ttgcctgggt ggagggtgag acaaagtatg tctccgatgc cactgtggcc 1080
atctttgtgg ccaccctgct attcattgtg ccttcacaga agcccaagtt taacttccgc 1140
agccagactg aggaagaaag gaaaactcca ttttatcccc ctcccctgct ggattggaag 1200
gtaacccagg agaaagtgcc ctggggcatc gtgctgctac taggggggcgg atttgctctg 1260
gctaaaggat ccgaggcctc ggggctgtcc gtgtggatgg ggaagcagat ggagcccttg 1320
cacgcagtgc ccccggcagc catcaccttg atcttgtcct tgctcgttgc cgtgttcact 1380
gagtgcacaa gcaacgtggc caccaccacc ttgttcctgc ccatctttgc ctccatgtct 1440
cgctccaacg gcctcaatcc gctgtacatc atgctgccct gtaccctgag tgcctccttt 1500
gccttcatgt tgcctgtggc caccctcca aatgccatcg tgttcaccta tgggcacctc 1560
aaggttgctg acatggtgaa aacaggagtc ataatgaaca taattggagt cttctgtgtg 1620
tttttggctg tcaacacctg gggacgggcc atatttgact tggatcattt ccctgactgg 1680
gctaatgtga cacatattga gacttaggaa gagccacaag accacacaca cagcccttac 1740
cctcctcagg actaccgaac cttctggcac accttgtaca gagttttggg gttcacaccc 1800
caaaatgacc caacgatgtc cacacaccac caaaacccag ccaatgggcc acctcttcct 1860
ccaagcccag atgcagagat ggtcatgggc agctggaggg taggctcaga aatgaaggga 1920

```
acccctcagt gggctgctgg acccatcttt cccaagcctt gccattatct ctgtgaggga 1980
ggccaggtag ccgagggatc aggatgcagg ctgctgtacc cgctctgcct caagcatccc 2040
ccacacaggg ctctggtttt cactcgcttc gtcctagata gtttaaatgg gaatcggatc 2100
ccctggttga gagctaagac aaccacctac cagtgcccat gtcccttcca gctcaccttg 2160
agcagcctca gatcatctct gccactctgg aagggacacc ccagccaggg acggaatgcc 2220
tggtcttgag caacctccca ctgctggagt gcgagtggga atcagagcct cctgaagcct 2280
ctgggaactc ctcctgtggc caccaccaaa ggatgaggaa tctgagttgc caacttcagg 2340
acgacacctg gcttgccacc cacagtgcac cacaggccaa cctacgccct tcatcacttg 2400
gttctgtttt aatcgactgg ccccctgtcc cacctctcca gtgagcctcc ttcaactcct 2460
tggtcccctg ttgtctgggt caacatttgc cgagacgcct tggctggcac cctctggggt 2520
ccccctttc tcccaggcag gtcatctttt ctgggagatg cttcccctgc catccccaaa 2580
tagctaggat cacactccaa gtatgggcag tgatggcgct ctgggggcca cagtgggcta 2640
tctaggccct ccctcacctg aggcccagag tggacacagc tgttaatttc cactggctat 2700
gccacttcag agtctttcat gccagcgttt gagctcctct gggtaaaatc ttccctttgt 2760
tgactggcct tcacagccat ggctggtgac aacagaggat cgttgagatt gagcagcgct 2820
tggtgatctc tcagcaaaca acccctgccc gtgggccaat ctacttgaag ttactcggac 2880
aaagacccca aagtggggca acaactccag agaggctgtg ggaatcttca gaacccccct 2940
gtaagagaca gacatgagag acaagcatct tctttccccc gcaagtccat tttatttcct 3000
tcttgtgctg ctctggaaga gaggcagtag caaagagatg agctcctgga tggcattttc 3060
cagggcagga gaaagtatga gagcctcagg aaaccccatc aaggaccgag tatgtgtctg 3120
gttccttggg tgggacgatt cctgaccaca ctgtccagct cttgctctca ttaaatgctc 3180
tgtctcccgc ggaaaaaaaa a                                          3201
```

<210> 44
<211> 3688
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 90113658CB1

<400> 44

```
gctgtgtcca agaaaggagt cagcaaaggg tgatgggatt atcattagtt cttataggtt 60
tgggataggc ggtggagtta ggagcaattt tttataggca gggggtggat ctcacaaagt 120
acattctcaa gggcggggag aatattacaa agtaccttct taagggcggg ggagtatgac 180
aaagtatatt attgcaaggg cggggagggg gtattgtcat aaggtcaatt gatcagttag 240
ggtgggtagg aacagatcac aatggtggaa tgttatttttt tgtggttctt catttgcttc 300
aggtcatctg gatgtatatg tgcaggtcac aggagatgtg atggcttagc ttgggctcag 360
aggcctgaca gtttgtatcc tcgtctccgt aatggtgcaa tgatccttcc ttcgcgtagg 420
gtaccatgtg tctcagggct ggaggaccac ctggccttgc tggcctctta caccagagcc 480
tgggccatag ctcataattt tgctgatatt ggcatctttt aacctcagaa agtatactct 540
gggtctaata atctgcctcc ttcaggaata ttcattcctc ttgttgcccg ctagcttcca 600
gtcataaaaa aaaatccgtt taataacttt agatttggtt acatccatct gtgtgttttt 660
cttcacctcc atttagaatt agtaggaaaa gcattagaac aaaagggttg tttaaagaac 720
attagtttat ctcatttgaa aagtgaaact tctttcatgt aagctctggt tacttcattg 780
ttctttccac aagcccttct aacaatgatt atctgcagtt tgcccaccca ctgtcaagag 840
ctcagcatcc atacttgaaa ctcaagcttc atatgccaac tttgggatga ttctcatcaa 900
acagaagagg ctgtttccat gctggatccc ggcgctgttc atcggcttca gccagttctc 960
ggactcgttc ctcctggacc agcccaactt ctggtgccgc ggggccggca aaggcaccga 1020
gctggcaggg gtcaccacca caggccgggg cggggacatg ggcaactgga ccagcctccc 1080
caccacccca ttcgccactg ccccctggga ggctgcgggc aaccggagca acagcagcgg 1140
cgcggacgga ggcgacacac caccctgcc atcccctccg gacaaggggg acaacgcctc 1200
caactgtgac tgccgcgcat gggactacgg catccgcgcc ggcctcgtcc agaacgtggt 1260
cagcaagtgg gatcttgtgt gtgataatgc ctggaaggtc catatcgcta agttctcctt 1320
actggttgga ttaatctttg gctacctaat aactggatgc attgctgact gggtcggccg 1380
```

gcggcctgtg ctgctgtttt ccatcatctt cattctgatc tttggactga ctgtggcact 1440
gtcagtgaat gtgacaatgt tcagcacact caggttcttt gaaggatttt gcctggctgg 1500
aatcattctc accttgtatg ctttacgaat agagctgtgc ccccctggaa aacggttcat 1560
gattacgatg gtggcgagct tcgtggccat ggcgggccag ttcctcatgc ctgggctagc 1620
cgccctgtgc cgggattggc aggtgctgca ggccctcatc atctgccctt tcctgctcat 1680
gctgctctac tggtcgatat tccccgagtc cctccggtgg ctaatggcca cccagcagtt 1740
tgagtctgca aagaggctga tcctccactt cacacagaag aatcgcatga accctgaggg 1800
cgacatcaag ggtgtgatac cagagctgga gaaagagctt tcccggaggc ccaagaaggt 1860
ctgcatcgtg aaggtggtgg ggacacggaa cctgtggaag aacattgtgg tcctgtgtgt 1920
gaactcgctg acggggtacg ggatccacca ctgctttgcc aggagcatga tgggccacga 1980
ggtgaaggtg ccgctcctgg agaacttcta tgctgactac tataccacgg ccagcatcgc 2040
gctggtgtcc tgcctggcca tgtgcgtggt ggtccgattc ctcgggcgca ggggagggct 2100
gctgctcttc atgatcctca ccgccctggc ctcgctcctg cagctcggcc tcctcaacct 2160
gattggaaag tacagccagc acccagactc agggatgagt gacagcgtca aggacaaatt 2220
ttccatcgcg ttttccatcg tgggcatgtt tgcctcccat gcggtgggga gcctcagcgt 2280
gttcttctgt gcggagatca ccccgacggt gataaggtgt ggcgggctgg ggctggtgct 2340
ggccagcgcg ggcttcggca tgctgacggc acccatcatc gagctgcaca accagaaagg 2400
ctacttcctg caccacatca tctttgcctg ctgcacgctc atctgcatca tctgcatcct 2460
cctgctgccc gagagcaggg accagaacct gcctgagaac atttctaacg gggagcacta 2520
cacgcgccag ccgctgctgc cgcacaagaa gggggagcag ccactgctgc tcaccaacgc 2580
cgagctcaag gactactcgg gcctccacga tgccgcagcc gcgggtgaca cactgcccga 2640
gggtgccacg gccaacggca tgaaggccat gtagcccggc ctgcggaacc cggggctcca 2700
gggtctgggg cagcttgggc acaggtttac agaccaggga ccgaacacgc agccaggggt 2760
gggaaagctg cctcagccaa gctgagcctc tcaactggtg tggggaaatc ctgtctttcc 2820
aaaagtccaa ggagcgcggg tcggaggaga caaactcttt ggaaataacc ctttcaagac 2880
tttcttttct gccgttaaat gtgtgtattt attttggtca tttttacgag aagcacttta 2940
ttccctctcc ctctcactga tcacaaatgg aatcacctcc ctgggcagcg agacgcagtt 3000
gctctgggaa gatgccacag tgaccagggc catggccggt ccctctgggg agatgggaca 3060
cggctcctgg cagcacccag gagcccccca ggctgcctgc ctgcgtgcag aggcaggaga 3120
ggaccgtgga gcgtttccgg gactgcattt tagacggagt gaaatgtaca tgaatttggc 3180
ttttgctaga gtctgtgtat ggtttttaa ggtccttttt ccctttctgt ttgtaaggta 3240
agagcttctg ttcgtgtgca gggaaagcag ctcacagacg ccgttaaaac cagcttcatc 3300
tttccttcag gatcatcctt ttgtacttga tgctggaagc tcttggagaa aaagcttaaa 3360
catttcacca gaaatcttaa ttgagcagca gtcatatcgc cacagctttg tgagtacaca 3420
gctaacagat gcgtcgagac ctgagatgtg ctcgtttttа ttcctcctct cccagattgg 3480
ctccagcaga aaagtccctc gtgcacaggc agctcttgtg ccggcactac ttgaaaacat 3540
ggctactttc taagctacaa accattagaa aataactcaa aaagatggca gaggcaaatc 3600
cacagaaggg gggctgccct ccacacacac acgcccgtca cccacacctt gagcacacac 3660
catgaagatc acctactgag gaggatcc 3688

<210> 45
<211> 2402
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3942766CB1

<400> 45
gcccgtaccg ccaggcgatc gcgctgatgg cggcgctggc agcagcggcc aagaaggtgt 60
ggagcgcgcg gcggctgctg gtgctgctgt tcacgccgct cgcgctgctg ccggtggtct 120
tcgccctccc gcccaaggaa ggccgctgct tgtttgtcat cctgctcatg gcggtgtact 180
ggtgcacgga ggccctgccg ctctcagtga cggcgctgct gcccatcgtc ctcttccccct 240
tcatgggcat cttgccctcc aacaaggtct gcccccagta cttcctcgac accaacttcc 300
tcttcctcag tgggctgatc atggccagcg ccattgagga gtggaacctg caccggcgaa 360

tcgccctcaa gatcctgatg cttgttggag tccagccggc caggctcatc ctggggatga 420
tggtgaccac ctcgttcttg tccatgtggc tgagcaacac cgcctccact gccatgatgc 480
ttcccattgc caatgccatc ctgaaaagtc tctttggcca gaaggaggtt cgaaaggacc 540
ccagccagga gagtgaagag aacacagctg ctgtgcggag aaacggccta cacactgtgc 600
ccacggagat gcagtttctc gccagcacag aagccaaaga ccaccctggg gagacagagg 660
ttccactgga tctgccggct gactccagga aggaggatga atatcgtcgg aacatctgga 720
agggcttcct catctccatc ccctactcag ccagtattgg gggcacagcc acactcacgg 780
gcacagcccc taacctcatc ctgcttggcc agctcaagag tttctttccg cagtgtgacg 840
tggtgaattt cggctcctgg ttcattttcg ccttccctct tatgctgttg ttcctgttgg 900
caggctggct ctggatctcc ttcctgtacg ggggactgag cttcaggggc tggaggaaga 960
ataaatctga gataagaacc aatgcagaag atagggctcg agctgtaatt cgggaagaat 1020
accagaacct ggggcccatc aagtttgccg aacaggctgt tttcatcctt ttctgcatgt 1080
ttgccatcct cctcttcacc cgggacccga agttcatccc tggctgggcc agcctcttca 1140
atcctgggtt tctttctgat gctgtcaccg gcgtggctat tgtcaccatc ttgttcttct 1200
tcccgtccca aaggccctct ctcaagtggt ggtttgactt caaagctccc aacacagaga 1260
cagagccctt gctgacctgg aagaaggccc aggagacagt gccctggaac atcatccttc 1320
tcctgggagg gggcttcgcc atggccaaag gctgtgagga atcggggctg tctgtatgga 1380
ttggtgggca gctgcacccc ctggagaatg tgccccccgc cctggctgtg ctgctcatca 1440
ctgtggtcat cgccttcttc actgagtttg ccagcaacac ggcgaccatc atcatcttcc 1500
tgccggtcct ggcagagctg gccatccgcc tgagagtgca ccccctgtat ctgatgattc 1560
cgggcacagt cggctgctcc tttgccttca tgctcccggt ctcaacgccc cccaactcca 1620
tcgccttcgc ctctggacac ttgctggtca aagacatggt gcggacaggc ctcctgatga 1680
acctgatggg tgtcctgctg ctcagtttgg ctatgaatac ctgggcacag accatcttcc 1740
agctgggcac cttcccggac tgggctgata tgtactcggt caatgtcaca gcattgccac 1800
ccaccttggc caatgacaca tttcggaccc tctgagtccc cctggaggac tcccttgagg 1860
ctgaaccctc tcaaagggct gtcaccatca cctgctgcta ggacactaca aaacaatcaa 1920
ataattcttt tctgtaatcc aatatgcagc aagcaagggt gacctccagt ggcccactca 1980
agtccatgag ccattatcta ggatactttc tctctctttc atgcagttca aagcccaggt 2040
atctctcaga tctgctgcct gagaaataag ctcctttatc agttagctgt tttatcatta 2100
ggatacaaga cagcccagtg tcatcaacag tgagcaaatc tggcatggtg ttgtctcgta 2160
cagtgggata ggaggccatt cattcccatg ggcacagcta acattatccc ccagtgatta 2220
ctttcgatta cactgaagag acacctgtct taaagatcac tttcctgggc tggcatgtcc 2280
aaccagttgt tccagcctgt tcacaccagg gggggattgc ttagttggtg cccatctgcc 2340
cagtgtttac cctgcgatac caggagcaga gatcctttaa tcctgggacc ggaaggcagt 2400
ac 2402

<210> 46
<211> 2410
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7501987CB1

<400> 46
catccgctca caatgccaca tcaatgatac gagcacgtag cctcactgct tgcacagtgc 60
atggcagagt cggctgcgag caggcgaggt ggcctgaggg aggtcactag gctggctgag 120
ggctttttgc tgtggttctg agccggcctg cttccaggca ccgtgtccat gcgggtaagc 180
ggtctccctg ggtgcccact cttgcgcccg gagatcctga gtttggtcct gtctggccat 240
gaagctcagc ctgctgggag gccacaggga gatgcaggct gggcggcggg tggatggttc 300
cagccggttg ggtccggggc ctggagctca gcctgtgggg tggggaccca gtggtgccct 360
ggagctgccg cttctgctct cagcaggatg atgggcagga cagggagagg ctgacctact 420
tccagaacct gcctgagtct ctgacttccc tcctggtgct gctgaccacg gccaacaacc 480
ccgatgtgat gattcctgcg tattccaaga accgggccta tgccatcttc ttcatagtct 540
tcactgtgat aggaagcctg tttctgatga acctgctgac agccatcatc tacagtcagt 600

tccggggcta cctgatgaaa tctctccaga cctcgctgtt tcggaggcgg ctgggaaccc 660
gggctgcctt tgaagtccta tcctccatgg tgggggaggg aggagccttc cctcaggcag 720
ttggggtgaa gccccagaac ttgctgcagg tgcttcagaa ggtccagctg gacagctccc 780
acaaacaggc catgatggag aaggtgcgtt cctatggcag tgttctgctc tcagctgagg 840
agtttcagaa gctcttcaac gagcttgaca gaagtgtggt taaagagcac ccgccgaggc 900
ccgagtacca gtctccgttt ctgcagagcg cccagttcct cttcggccac tactactttg 960
actacctggg gaacctcatc gccctggcaa acctggtgtc catttgcgtg ttcctggtgc 1020
tggatgcaga tgtgctgcct gctgagcgtg atgacttcat cctggggatt ctcaactgcg 1080
tcttcattgt gtactacctg ttggagttgc tgctcaaggt ctttgccctg ggcctgcgag 1140
ggtacctgtc ctaccccagc aacgtgtttg acgggctcct caccgttgtc ctgctggagg 1200
ccggagatgg tgggcctgct gtcgctgtgg gacatgaccc gcatgctgaa catgctcatc 1260
gtgttccgct tcctgcgtat catccccagc atgaagccga tggccgtggt ggccagtacc 1320
gtcctgggcc tggtgcagaa catgcgtgcg tttggcggga tcctggtggt ggtctactac 1380
gtatttgcca tcattgggat caacttgttt agaggcgtca ttgtggctct tcctggaaac 1440
agcagcctgg cccctgccaa tggctcggcg ccctgtggga gcttcgagca gctggagtac 1500
tgggccaaca acttcgatga ctttgcggct gccctggtca ctctgtggaa cttgatggtg 1560
gtgaacaact ggcaggtgtt tctggatgca tatcggcgct actcaggccc gtggtccaag 1620
atctattttg tgttgtggtg gctggtgtcg tctgtcatct gggtcaacct gtttctggcc 1680
ctgattctgg agaacttcct tcacaagtgg gacccccgca gccacctgca gccccttgct 1740
gggaccccag aggccaccta ccagatgact gtggagctcc tgttcaggga tattctggag 1800
gagcccgggg aggatgagct cacagagagg ctgagccagc acccgcacct gtggctgtgc 1860
aggtgacgtc cgggctgccg tcccagcagg ggcggcagga gagagaggct ggcctacaca 1920
ggtgcccatc atggaagagg cggccatgct gtggccagcc aggcaggaag agacctttcc 1980
tctgacggac cactaagctg gggacaggaa ccaagtcctt tgcgtgtggc ccaacaacca 2040
tctacagaac agctgctggt gcttcaggga ggcgccgtgc cctccgcttt cttttatagc 2100
tgcttcagtg agaattccct cgtcgactcc acagggacct ttcagacaaa aatgcaagaa 2160
gcagcggcct cccctgtccc ctgcagcttc cgtggtgcct ttgctgccgg cagcccttgg 2220
ggaccacagg cctgaccagg gcctgcacag gttaaccgtc agacttccgg ggcattcagg 2280
tggggatgct ggtggtttga catggagaga accttgactg tgttttatta tttcatggct 2340
tgtatgagtg tgactgggtg tgtttcttta gggttctgat tgccagttat tttcatcaat 2400
aagtcttgca 2410

<210> 47
<211> 968
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503223CB1

<400> 47
gaggagatgg gttccaattt tagggagcag gctgcctaat gaaggagcca ggcttgcaca 60
cagacaattc tagaactggt ggcccgagag ggatgtgaag gcccaaaatg accctcttac 120
cgggagacaa ttctgactac gactacagcg cgctgagctg cacctcggac gcctccttcc 180
acccggcctt cctcccgcag cgccaggcca tcaagggcgc gttctaccgc cgggcgcagc 240
ggctgcggcc gcaggatgag ccccgccagg gctgtcagcc cgaggaccgc cgccgtcgga 300
tcatcatcaa cgtaggcggc atcaagtact cgctgccctg gaccacgctg gacgagttcc 360
cgctgacgcg cctgggccag ctcaaggcct gcaccaactt cgacgacatc ctcaacgtgt 420
gcgatgacta cgacgtcacc tgcaacgagt tcttcttcga ccgcaacccg ggggccttcg 480
gcactatcct gaccttcctg cgcgcgggca agctgcggct gctgcgcgag atgtgcgcgc 540
tgtccttcca ggacagtgac atcttgttcg gaagtgcctc ctcggacacc agagacaata 600
actgagcgcg gaggacacgc ctgccctgcc tgccatctgt ggcccgaagc cattgccatc 660
cactgcagac gcctggagag ggacaggccg cttccgagtg cagtcctggc gcagcaccga 720
ctcccacgca cccggggaag gacaccctca ctcccacacc ccgggaagaa cactagaaca 780
tcagcagagg ggccctgccc ctccgcctgc agccgtgaaa ggaagctggg tcatcagccc 840

agccccgccc accccagccc ctatgtgtgt ttccctcaat aaggagatgc cttgttcttt 900
tcaccatgca aataacatgc ccagcaaaaa cttgctttat gggtctgcct ggagaaaaaa 960
aaaaaaaa 968

<210> 48
<211> 2267
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503566CB1

<400> 48
gtcttgctaa cagctgccaa tacctcactg agtgcctcac accaacatgg gctccaagtg 60
agtttccttc gtctgggcag actccctccc ctcttccata aaggctgcag gagacctgta 120
gctgtcacag gaccttccct aagagcccgc aggggaagac tgccccagtc cggccatcac 180
catgctccgg accattctgg atgctcccca gcggttgctg aaggagggga gagcgtcccg 240
gcagctggtg ctggtggtgg tattcgtcgc tttgctcctg gacaacatgc tgtttactgt 300
ggtggtgcca attgtgccca ccttcctata tgacatggag ttcaaagaag tcaactcttc 360
tctgcacctc ggccatgccg gaagttcccc acatgccctc gcctctcctg ccttttccac 420
catcttctcc ttcttcaaca acaacaccgt ggctgttgaa gaaagcgtac ctagtggaat 480
agcatggatg aatgacactg ccagcaccat cccacctcca gccactgaag ccatctcagc 540
tcataaaaac aactgcttgc aaggcacagg tttcttggag gaagagacta cccgggtcgg 600
ggttctgttt gcttcaaagg ctgtgatgca acttctggtc aacccattcg tgggccctct 660
caccaacagg attggatatc atatccccat gtttgctggc tttgttatca tgtttctctc 720
cacagttagt cttggaatgc tggccagtgt ctacactgat gaccatgaga gaggacgagc 780
catgggaact gctctggggg gcctggcctt ggggttgctg gtgggagctc cctttggaag 840
tgtaatgtac gagtttgttg ggaagtctgc acccttcctc atcctggcct tcctggcact 900
actggatgga gcactccagc tttgcatcct acagccttcc aaagtctctc ctgagagtgc 960
caaggggact cccctcttta tgcttctcaa agacccttac atcctggtgg ctgcagggtc 1020
catctgcttt gccaacatgg gggtggccat cctggagccc acactgccca tctggatgat 1080
gcagaccatg tgctccccca agtggcagct gggtctagct ttcttgcctg ccagtgtgtc 1140
ctacctcatt ggcaccaacc tctttggtgt gttggccaac aagatgggtc ggtggctgtg 1200
ttccctaatc gggatgctgg tagtaggtac cagcttgctc tgtgttcctc tggctcacaa 1260
tatttttggt ctcattggcc ccaatgcagg gcttggcctt gccataggca tggtggattc 1320
ttctatgatg cccatcatgg ggcacctggt ggatctacgc cacacctcgg tgtatgggag 1380
tgtctacgcc atcgctgatg tggctttttg catgggcttt gctataggtc catccaccgg 1440
tggtgccatt gtaaaggcca tcggttttcc ctggctcatg gtcatcactg gggtcatcaa 1500
catcgtctat gctccactct gctactacct gcggagcccc ccggcaaagg aagagaagct 1560
tgctattctg agtcaggact gccccatgga gacccggatg tatgcaaccc agaagcccac 1620
gaaggaattt cctctggggg aggacagtga tgaggagcct gaccatgagg agtagcagca 1680
gaaggtgctc cttgaattca tgatgcctca gtgaccacct ctttccctgg gaccagatca 1740
ccatggctga gcccacggct cagtgggctt cacatacctc tgcctgggaa tcttctttcc 1800
tcccctccca tggacactgt ccctgatact cttctcacct gtgtaacttg tagctcttcc 1860
tctatgcctt ggtgccgcag tggcccatct tttatgggaa gacagagtga tgcaccttcc 1920
cgctgctgtg aggttgatta aacttgagct gtgacgggtt ctgcaagggg tgactcattg 1980
catagaggtg gtagtgagta atgtgccccc gaaaccagtg gggtgactga caagcctctt 2040
taatctgttg cctgattttc tctggcatag tcccaacaga tcggaagagt gttaccctct 2100
tttcctcaac gtgttctttc ccgggttttc ccagccgagt tgagaaaatg ttctcagcat 2160
tgtcttgctg ccaaatgcca gcttgaagag ttttgttttg ttttttttcc atttattttt 2220
ttttttaata aagtgagtga tttttctgtg gctaaaaaaa aaaaaaa 2267

<210> 49
<211> 319
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7505122CB1

<400> 49
cggctcgagc tcgaccgaat cggctcgagc ggctcgagtg ctcagcctgg tgaaccacac 60
aggccagcgc tctgacatgc agaaggtgac cctgggcctg cttgtgttcc tggcaggctt 120
tcctgtcctg gacgccaatg acctagaaga taaaaacagt cctttctact atgactggca 180
cagcctccag gttggcgggc tcatctgcgc tggggttctg tgcatggcag ggcctcatct 240
cacctctcgc aagagggtct ctttgttcaa ttttttttaa tctaaaatga ttgtgcctct 300
gcccaaaaaa aaaaaaaaa 319

<210> 50
<211> 2510
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511620CB1

<400> 50
ggagcagccc gcaccggaca acttgcgagc catggggctg gcggatgcgt cgggaccgag 60
ggacacacag gcactgctgt ctgcaacaca agcaatggac ctgcggaggc gagactacca 120
catggaacgg ccgctgctga accaggagca tttggaggag ctggggcgct ggggctcagc 180
acctaggacc caccagtggc ggacctggtt gcagtgctcc cgtgctcggg cctatgccct 240
tctgctccaa cacctcccgg ttttggtctg gttaccccgg tatcctgtgc gtgactggct 300
cctgggtgac ctgttatccg gcctgagtgt ggccatcatg cagcttccgc agggcttggc 360
ctacgccctc ctggctggat tgccccccgt gtttggcctc tatagctcct tctaccctgt 420
cttcatctac ttcctgtttg gcacttcccg gcacatctcc gtgggccttg aacgactcca 480
tgatcaatga gacagccaga gatgctgccc gggtacaggt ggcctccaca ctcagtgtcc 540
tggttggcct cttccaggtg gggctgggcc tgatccactt cggcttcgtg gtcacctacc 600
tgtcagaacc tcttgtccga ggctatacca cagctgcagc tgtgcaggtc ttcgtctcac 660
agctcaagta tgtgtttggc ctccatctga gcagccactc tgggccactg tccctcatct 720
atacagtgct ggaggtctgc tggaagctgc cccagagcaa ggttggcacc gtggtcactg 780
cagctgtggc tggggtggtg ctcgtggtgg tgaagctgtt gaatgacaag ctgcagcagc 840
agctgcccat gccgataccc ggggagctgc tcacgctcat cggggccaca ggcatctcct 900
atggcatggg tctaaagcac agatttgagg tagatgtcgt gggcaacatc cctgcagggc 960
tggtgccccc agtggccccc aacacccagc tgttctcaaa gctcgtgggc agcgccttca 1020
ccatcgctgt ggttgggttt gccattgcca tctcactggg gaagatcttc gccctgaggc 1080
acggctaccg ggtggacagc aaccaggagc tggtggccct gggcctcagt aaccttatcg 1140
gaggcatctt ccagtgcttc cccgtgagtt gctctatgtc tcggagcctg gtacaggaga 1200
gcaccggggg caactcgcag gttgctggag ccatctcttc ccttttcatc ctcctcatca 1260
ttgtcaaact tggggaactc ttccatgacc tgcccaaggc ggtcctggca gccatcatca 1320
ttgtgaacct gaagggcatg ctgaggcagc tcagcgacat gcgctccctc tggaaggcca 1380
atcgggcgga tctgcttatc tggctggtga ccttcacggc caccatcttg ctgaacctgg 1440
accttggctt ggtggttgcg gtcatcttct ccctgctgct cgtggtggtc cggacacaga 1500
tgccccacta ctctgtcctg gggcaggtgc cagacacgga tatttacaga gatgtggcag 1560
agtactcaga ggccaaggaa gtccgggggg tgaaggtctt ccgctcctcg gccaccgtgt 1620
actttgccaa tgctgagttc tacagtgatg cgctgaagca gaggtgtggt gtggatgtcg 1680
acttcctcat ctcccagaag aagaaactgc tcaagaagca ggagcagctg aagctgaagc 1740
aactgcagaa agaggagaag cttcggaaac aggctgcctc ccccaagggc gcctcagttt 1800
ccattaatgt caacaccagc cttgaagaca tgaggagcaa caacgttgag gactgcaaga 1860
tgatggtgag ctcaggagat aagatggaag atgcaacagc caatggtcaa gaagactcca 1920

aggccccaga tgggtccaca ctgaaggccc tgggcctgcc tcagccagac ttccacagcc 1980
tcatcctgga cctgggtgcc ctctcctttg tggacactgt gtgcctcaag agcctgaaga 2040
atattttcca tgacttccgg gagattgagg tggaggtgta catggcggcc tgccacagcc 2100
ctgtggtcag ccagcttgag gctgggcact tcttcgatgc atccatcacc aagaagcatc 2160
tctttgcctc tgtccatgat gctgtcacct ttgccctcca acacccgagg cctgtccccg 2220
acagccctgt ttcggtcacc agactctgaa catgctacat cctgcccaag actgcacctc 2280
tggaggtgca gggcaccctt gagaagcccc tcacccctag gccgcctcca ggtgctaccc 2340
aggagtcccc tccatgtaca cacacacaac tcagggaagg aggtcctggg actccaagtt 2400
cagcgctcca ggtctgggac agggcctgca tgcagtcagg ctggcagtgg cgcggtacag 2460
ggagggaact ggtgcatatt ttagcctcag gaataaagat ttgtctgctc 2510

<210> 51
<211> 2241
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506995CB1

<400> 51
catccgctca caatgccaca tcaatgatac gagcacgtag cctcactgct tgcacagtgc 60
atggcagagt cggctgcgag caggcgaggt ggcctgaggg aggtcactag gctggctgag 120
ggctttttgc tgtggttctg agccggcctg cttccaggca ccgtgtccat gcgggtgagc 180
ggtctccctg ggtgcccact cttgcgcccg gagatcctga gtttggtcct gtctggccat 240
gaagctcagc ctgctgggag gccacaggga gatgcaggct gggcggcggg tggatggttc 300
cagccggttg ggtccggggc ctggagctca gcctgtgggg tggggaccca gtggtgccct 360
ggagctgccg cttctgctct cagcaggatg atgggcagga cagggagagg ctgacctact 420
tccagaacct gcctgagtct ctgacttccc tcctggtgct gctgaccacg gccaacaacc 480
ccgatgtgat gattcctgcg tattccaaga accgggccta tgccatcttc ttcatagtct 540
tcactgtgat aggaagcctg tttctgatga acctgctgac agccatcatc tacagtcagt 600
tccggggcta cctgatgaaa tctctccaga cctcgctgtt tcggaggcgg ctgggaaccc 660
gggctgcctt tgaagtccta tcctccatgg tgggggaggg aggagccttc cctcaggcca 720
cccgccgagg cccgagtacc agtctccgtt tctgcagagc gcccagttcc tcttcggcca 780
ctactacttt gactacctgg ggaacctcat cgccctggca aacctggtgt ccatttgcgt 840
gttcctggtg ctggatgcag atgtgctgcc tgctgagcgt gatgacttca tcctggggat 900
tctcaactgc gtcttcattg tgtactacct gttggagatg ctgctcaagg tctttgccct 960
gggcctgcga gggtacctgt cctaccccag caacgtgttt gacgggctcc tcaccgttgt 1020
cctgctggag gccggagatg gtgggcctgc tgtcgctgtg ggacatgacc cgcatgctga 1080
acatgctcat cgtgttccgc ttcctgcgta tcatccccag catgaagccg atggccgtgg 1140
tggccagtac cgtcctgggc ctggtgcaga acatgcgtgc ttttggcggg atcctggtgg 1200
tggtctacta cgtatttgcc atcattggga tcaacttgtt tagaggcgtc attgtggctc 1260
ttcctggaaa cagcagcctg gcccctgcca atggctcggc gccctgtggg agcttcgagc 1320
agctggagta ctgggccaac aacttcgatg actttgcggc tgccctggtc actctgtgga 1380
acttgatggt ggtgaacaac tggcaggtgt ttctggatgc atatcggcgc tactcaggcc 1440
cgtggtccaa gatctatttt gtgttgtggt ggctggtgtc gtctgtcatc tgggtcaacc 1500
tgtttctggc cctgattctg gagaacttcc ttcacaagtg ggacccccgc agccacctgc 1560
agccccttgc tgggacccca gaggccacct accagatgac tgtggagctc ctgttcaggg 1620
atattctgga ggagcccggg gaggatgagc tcacagagag gctgagccag cacccgcacc 1680
tgtggctgtg caggtgacgt ccgggctgcc gtcccagcag gggcggcagg agagagaggc 1740
tggcctacac aggtgcccat catggaagag gcggccatgc tgtggccagc caggcaggaa 1800
gagacctttc ctctgacgga ccactaagct ggggacagga accaagtcct ttgcgtgtgg 1860
cccaacaacc atctacagaa cagctgctgg tgcttcaggg aggcgccgtg ccctccgctt 1920
tcttttatag ctgcttcagt gagaattccc tcgtcgactc cacagggacc tttcagacaa 1980
aaatgcaaga agcagcggcc tcccctgtcc cctgcagctt ccgtggtgcc tttgctgccg 2040
gcagcccttg gggaccacag gcctgaccag ggcctgcaca ggttaaccgt cagacttccg 2100

gggcattcag gtggggatgc tggtggtttg acatggagag aaccttgact gtgttttatt 2160
atttcatggc ttgtatgagt gtgactgggt gtgtttcttt agggttctga ttgccagtta 2220
ttttcatcaa taagtcttgc a 2241

<210> 52
<211> 2312
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506996CB1

<400> 52
catccgctca caatgccaca tcaatgatac gagcacgtag cctcactgct tgcacagtgc 60
atggcagagt cggctgcaga gcaggcgagg tggcctgagg gaggtcacta ggctggctga 120
gggctttttg ctgtggttca tgagccggcc tgcttccagg caccgtgtcc atgcgggtga 180
gcggtctccc tgggtgccca ctcttgcgcc cggagatcct gagtttggtc ctgtctggcc 240
atgaagctca gcctgctggg aggccacagg gagatgcagg ctgggcggcg ggtggatggt 300
tccagccggt tgggtccggg gcctggagct cagcctgtgg ggtggggacc cagtggtgcc 360
ctggagctgc cgcttctgct ctcagcagga tgatgggcag gacagggaga ggctgaccta 420
cttccagaac ctgcctgagt ctctgacttc cctcctggtg ctgctgacca cggccaacaa 480
ccccgatgtg atgattcctg cgtattccaa gaaccgggcc tatgccatct tcttcatagt 540
cttcactgtg ataggaagcc tgtttctgat gaacctgctg acagccatca tctacagtca 600
gttccggggc tacctgatga aatctctcca gacctcgctg tttcggaggc ggctgggaac 660
ccgggctgcc tttgaagtcc tatcctccat ggtgggggag ggaggagcct tccctcaggc 720
agttggggtg aagccccaga acttgctgca ggtgcttcag aaggtccagc tggacagctc 780
ccacaaacag gccatgatgg agaaggtgcg ttcctatggc agtgttctgc tctcagctga 840
ggagtttcag aagctcttca acgagcttga cagaagtgtg gttaaagagc acccgccgag 900
gcccgagtac cagtctccgt ttctgcagag cgcccagttc ctcttcggcc actactactt 960
tgactacctg gggaacctca tcgccctggc aaacctggtg tccatttgcg tgttcctggt 1020
gctggatgca gatgtgctgc ctgctgagcg tgatgacttc atcctgggga ttctcaactg 1080
cgtcttcatt gtgtactacc tgttggagtt gctgctcaag gtctttgccc tgggcctgcg 1140
agggtacctg tcctacccca gcaacgtgtt tgacgggctc ctcaccgttg tcctgctgcc 1200
gatggccgtg gtggccagta ccgtcctggg cctggtgcag aacatgcgtg cttttggcgg 1260
gatcctggtg gtggtctact acgtatttgc catcattggg atcaacttgt ttagaggcgt 1320
cattgtggct cttcctggaa acagcagcct ggcccctgcc aatggctcgg cgccctgtgg 1380
gagcttcgag cagctggagt actgggccaa caacttcgat gactttgcgg ctgccctggt 1440
cactctgtgg aacttgatgg tggtgaacaa ctggcaggtg tttctggatg catatcggcg 1500
ctactcaggc ccgtggtcca agatctattt tgtgttgtgg tggctggtgt cgtctgtcat 1560
ctgggtcaac ctgtttctgg ccctgattct ggagaacttc cttcacaagt gggacccccg 1620
cagccacctg cagccccttg ctgggacccc agaggccacc taccagatga ctgtggagct 1680
cctgttcagg gatattctgg aggagcccgg ggaggatgag ctcacagaga ggctgagcca 1740
gcacccgcac ctgtggctgt gcaggtgacg tccgggctgc cgtcccagca ggggcggcag 1800
gagagagagg ctggcctaca caggtgccca tcatggaaga ggcggccatg ctgtggccag 1860
ccaggcagga agagaccttt cctctgacgg accactaagc tggggacagg aaccaagtcc 1920
tttgcgtgtg gcccaacaac catctacaga acagctgctg gtgcttcagg gaggcgccgt 1980
gccctccgct ttcttttata gctgcttcag tgagaattcc ctcgtcgact ccacagggac 2040
ctttcagaca aaaatgcaag aagcagcggc ctcccctgtc ccctgcagct tccgtggtgc 2100
ctttgctgcc ggcagccctt ggggaccaca ggcctgacca gggcctgcac aggttaaccg 2160
tcagacttcc ggggcattca ggtggggatg ctggtggttt gacatggaga gaaccttgac 2220
tgtgttttat tatttcatgg cttgtatgag tgtgactggg tgtgtttctt tagggttctg 2280
attgccagtt attttcatca ataagtcttg ca 2312